(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 074 617 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
07.02.2001 Bulletin 2001/06

(21) Application number: 00116126.4

(22) Date of filing: 28.07.2000

(51) Int Cl.7: C12N 15/12, C12N 15/11, C07K 14/47, C07K 16/18, C12Q 1/68

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: 29.07.1999 JP 24803699
27.08.1999 JP 30025399
11.01.2000 JP 2000118776
02.05.2000 JP 2000183767
09.06.2000 JP 2000241899

(71) Applicant: Helix Research Institute
Kisarazu-shi, Chiba 292-0812 JP

(72) Inventors:
- Ota, Toshio
  Fujisawa-shi, Kanagawa 251-0042 JP
- Isogai, Takao
  Inashiki-gun, Ibaraki 300-0303 JP
- Nishikawa, Tetsuo
  Itabashi-ku, Tokyo 173-0013 JP
- Hayashi, Kohji
  Ichihara-shi, Chiba 292-0056 JP
- Saito, Kaoru
  Kisarazu-shi, Chiba 292-0056 JP
- Yamamoto, Junichi
  Kisarazu-shi, Chiba 292-0041 JP
- Ishii, Shizuko
  Kisarazu-shi, Chiba 292-0812 JP
- Sugiyama, Tomoyasu
  Kisarazu-shi, Chiba 292-0045 JP
- Wakamatsu, Ai
  Kisarazu-shi, Chiba 292-0014 JP
- Nagai, Keiichi
  Higashiyamato-shi, Tokyo 207-0022 JP
- Otsuki, Tetsuji
  Kisarazu-shi, Chiba 292-0045 JP

(74) Representative: VOSSIUS & PARTNER
Siebertstrasse 4
81675 München DE

(54) Primers for synthesising full-length cDNA and their use

(57) Primers for synthesizing full-length cDNAs and their use are provided.

5602 cDNA encoding a human protein has been isolated and nucleotide sequences of 5'-, and 3' -ends of the cDNA have been determined. Furthermore, primers for synthesizing the full-length cDNA have been provided to clarify the function of the protein encoded by the cDNA. The full-length cDNA of the present invention containing the translation start site provides information useful for analyzing the functions of the protein.

EP 1 074 617 A2

Description

FIELD OF THE INVENTION

[0001] The present invention relates to a polynucleotide encoding a novel protein, a protein encoded by the polynucleotide, and new uses of these.

BACKGROUND OF THE INVENTION

[0002] Currently, the sequencing projects, the determination and analysis of the genomic DNA of various living organisms have been in progress all over the world. The whole genomic sequences of more than 10 species of prokaryotes, a lower eukaryote, yeast, and a multicellular eukaryote, C. elegans are already determined. As to human genome, which is supposed to be composed of three thousand million base pairs, the world wide cooperative projects have been under way to analyze it, and the whole structure is predicted to be determined by the years 2002-2003. The aim of the determination of genomic sequence is to reveal the functions of all genes and their regulation and to understand living organisms as a network of interactions between genes, proteins, cells or individuals through deducing the information in a genome, which is a blueprint of the highly complicated living organisms. To understand living organisms by utilizing the genomic information from various species is not only important as an academie subject, but also socially significant from the viewpoint of industrial application.

[0003] However, determination of genomic sequences itself cannot identify the functions of all genes. For example, as for yeast, only the function of approximately half of the 6000 genes, which is predicted based on the genomic sequence, was able to be deduced. As for human, the number of the genes is predicted to be approximately one hundred thousand. Therefore, it is desirable to establish "a high throughput analysis system of the gene functions" which allows us to identify rapidly and efficiently the functions of vast amounts of the genes obtained by the genomic sequencing.

[0004] Many genes in the eukaryotic genome are split by introns into multiple exons. Thus, it is difficult to predict correctly the structure of encoded protein solely based on genomic information. In contrast, cDNA, which is produced from mRNA that lacks introns, encodes a protein as a single continuous amino acid sequence and allows us to identify the primary structure of the protein easily. In human cDNA research, to date, more than one million ESTs (Expression Sequence Tags) are publicly available, and the ESTs presumably cover not less than 80% of all human genes.

[0005] The information of ESTs is utilized for analyzing the structure of human genome, or for predicting the exon-regions of genomic sequences or their expression profile. However, many human ESTs have been derived from proximal regions to the 3'-end of cDNA, and information around the 5'-end of mRNA is extremely little. Among these human cDNAs, the number of the corresponding mRNAs whose encoding protein sequences are deduced is approximately 7000, and further, the number of full-length therein is only 5500. Thus, even including cDNA registered as EST, the percentage of human cDNA obtained so far is estimated to be 10-15% of all the genes.

[0006] It is possible to identify the transcription start site of mRNA on the genomic sequence based on the 5'-end sequence of a full-length cDNA, and to analyze factors involved in the stability of mRNA that is contained in the cDNA, or in its regulation of expression at the translation stage. Also, since a full-length cDNA contains ATG, the translation start site, in the 5'-region, it can be translated into a protein in a correct frame. Therefore, it is possible to produce a large amount of the protein encoded by the cDNA or to analyze biological activity of the expressed protein by utilizing an appropriate expression system. Thus, analysis of a full-length cDNA provides valuable information which complements the information from genome sequencing. Also, full-length cDNA clones that can be expressed are extremely valuable in empirical analysis of gene function and in industrial application.

[0007] Therefore, if a novel human full-length cDNA is isolated, it can be used for developing medicines for diseases in which the gene is involved. The protein encoded by the gene can be used as a drug by itself. Thus, it has great significance to obtain a full-length cDNA encoding a novel human protein.

[0008] In particular, human secretory proteins or membrane proteins would be useful by itself as a medicine like tissue plasminogen activator (TPA), or as a target of medicines like membrane receptors. In addition, genes for signal transduction-associated proteins (protein kinases, etc.), glycoprotein-associated proteins, transcription-associated proteins, etc. are genes whose relationships to human diseases have been elucidated. Moreover, genes for disease-associated proteins form a gene group rich in genes whose relationships to human diseases have been elucidated.

[0009] Therefore, it has great significance to isolate novel full-length cDNA clones of human, only few of which has been isolated. Especially, isolation of a novel cDNA clone encoding a secretory protein or membrane protein is desired since the protein itself would be useful as a medicine, and also the clones potentially include a gene associated with diseases. In addition, genes encoding proteins that are associated with signal transduction, glycoprotein, transcription, or diseases are expected to be useful as target molecules for therapy, or as medicines themselves. These genes form a gene group predicted to be strongly associated with diseases. Thus, identification of the full-length cDNA clones

encoding those proteins has great significance.

SUMMARY OF THE INVENTION

[0010] An objective of the present invention is to provide a polynucleotide encoding a novel protein, a protein encoded by said polynucleotide, and novel usages of these.

[0011] The inventors have developed a method for efficiently cloning a human full-length cDNA that is predicted by the ATGpr etc. to be a full-length cDNA clone, from a full-length-enriched cDNA library that is synthesized by the oligo-capping method. Then, the inventors determined the nucleotide sequence of the obtained cDNA clones from both 5'- and 3'- ends.

[0012] Furthermore, the inventors analyzed the obtained clones by the BLAST search of the databases, SwissProt (http://www.ebi.ac.uk/ebi_docsSwissProt_db/swisshome.html), GenBank (http://www.ncbi.nlm.nih.gov/web/GenBank), and UniGene (Human) (http://www.ncbi.nlm.nih.gov/UniGene).

[0013] The full-length cDNA clones of the present invention have high fullness ratio since these were obtained by the combination of (1) construction of a full-length-enriched cDNA library that is synthesized by the oligo-capping method, and (2) a system in which the full-length ratio is evaluated from the nucleotide sequence of the 5'-end (selection based on the ATGpr, previously removed complete sequences to ESTs). However, the primer of the present invention enables to obtain full-length cDNA easily without any specialized methods as in the described method.

Homology analysis in which the analysis is carried out against a not-full-length cDNA fragment to postulate the function of a protein encoded by said fragment, is being commonly performed.

However, since such analysis is based on the information of the fragment, it is not clear as to whether this fragment corresponds to a part that is functionally important in the protein. In other words, the reliability of the homology analysis based on the information of a fragment is doubtful, as information related to the structure of the whole protein is not available. However, the homology analysis of the present invention is conducted based on the information of a full-length cDNA comprising the whole coding region of the cDNA, and therefore, the homology of various portions of the protein can be analyzed. Hence, the reliability of the homology analysis has been dramatically improved in the present invention.

[0014] The inventors completed the invention by finding that it is possible to synthesize a novel full-length cDNA by using the combination of a primer that is designed based on the nucleotide sequence of the 5'-ends of the selected full-length cDNA clones and any of an oligo-dT primer or a 3'-primer that is designed based on the nucleotide sequence of the 3'-ends of the selected clones.

[0015] Thus, the present invention relates to primers described below, a method for synthesizing a polynucleotide using the primers, and polynucleotides obtained by the method.

[0016] First, the present invention relates to

(1) use of an oligonucleotide as a primer for synthesizing the polynucleotide comprising the nucleotide sequence set forth in any one of SEQ ID NOs: 1-5547 and SEQ ID NOs: 16111-16164, or the complementary strand thereof, wherein said oligonucleotide is complementary to said polynucleotide or the complementary strand thereof and comprises at least 15 nucleotides;

(2) a primer set for synthesizing polynucleotides, the primer set comprising an oligo-dT primer and an oligonucleotide complementary to the complementary strand of the polynucleotide comprising the nucleotide sequence set forth in any one of SEQ ID NOs: 1-5547 and SEQ ID NOs: 16111-16164, wherein said oligonucleotide comprises at least 15 nucleotides; and

(3) a primer set for synthesizing polynucleotides, the primer set comprising a combination of an oligonucleotide comprising a nucleotide sequence complementary to the complementary strand of the polynucleotide comprising a 5'-end nucleotide sequence and an oligonucleotide comprising a nucleotide sequence complementary to the polynucleotide comprising a 3'-end nucleotide sequence, wherein said oligonucleotides comprise at least 15 nucleotides and wherein said combination of 5'-end nucleotide sequence / 3'-end nucleotide sequence is selected from the combinations of 5'-end nucleotide sequence / 3'-end nucleotide sequence set forth in the SEQ ID NOs in Tables 1 and 2.

[0017] Tables 1 and 2 shows names of clones obtained in the examples described later, comprising the polynucleotide of the present invention (Table 1, 5547 clones, Table 2; 54 clones), names of nucleotide sequences at the 5'-end and 3'-end of the full-length cDNA, and their corresponding SEQ ID NOs. A blank indicates that the 3'-end sequence corresponding to the 5'-end sequence has not been determined for the same clone.

[0018] The SEQ ID NO of a 5'-end sequence is shown on the right side of the name of the 5'-end sequence, and the SEQ ID NO of a 3'-end sequence is shown on the right side of the name of the 3'-end sequence.

NT2RP2001214	F-NT2RP2001214	16136	R-NT2RP2001214	16190
NT2RP2001460	F-NT2RP2001460	16137	R-NT2RP2001460	16191
NT2RP2001756	F-NT2RP2001756	16138	R-NT2RP2001756	16192
NT2RP2002056	F-NT2RP2002056	16139	R-NT2RP2002056	16193
NT2RP2002677	F-NT2RP2002677	16140	R-NT2RP2002677	16194
NT2RP2002755	F-NT2RP2002755	16141	R-NT2RP2002755	16195
NT2RP2002843	F-NT2RP2002843	16142	R-NT2RP2002843	16196
NT2RP2003101	F-NT2RP2003101	16143	R-NT2RP2003101	16197
NT2RP2003799	F-NT2RP2003799	16144	R-NT2RP2003799	16198
NT2RP2004095	F-NT2RP2004095	16145	R-NT2RP2004095	16199
NT2RP2004732	F-NT2RP2004732	16146	R-NT2RP2004732	16200
NT2RP2004920	F-NT2RP2004920	16147	R-NT2RP2004920	16201
NT2RP2005454	F-NT2RP2005454	16148	R-NT2RP2005454	16202
NT2RP2005776	F-NT2RP2005776	16149	R-NT2RP2005776	16203
NT2RP2005806	F-NT2RP2005806	16150	R-NT2RP2005806	16204
NT2RP2005882	F-NT2RP2005882	16151	R-NT2RP2005882	16205
NT2RP3001282	F-NT2RP3001282	16152	R-NT2RP3001282	16206
NT2RP3001723	F-NT2RP3001723	16153	R-NT2RP3001723	16207
NT2RP3002099	F-NT2RP3002099	16154	R-NT2RP3002099	16208
NT2RP3003155	F-NT2RP3003155	16155	R-NT2RP3003155	16209
NT2RP3004028	F-NT2RP3004028	16156	R-NT2RP3004028	16210
OVARC1000008	F-OVARC1000008	16157	R-OVARC1000008	16211
OVARC1000724	F-OVARC1000724	16158	R-OVARC1000724	16212
OVARC1000751	F-OVARC1000751	16159	R-OVARC1000751	16213
OVARC1001029	F-OVARC1001029	16160	R-OVARC1001029	16214
PLACE1000814	F-PLACE1000814	16161	R-PLACE1000814	16215
PLACE1003030	F-PLACE1003030	16162	R-PLACE1003030	16216
PLACE1005549	F-PLACE1005549	16163	R-PLACE1005549	16217
PLACE1007218	F-PLACE1007218	16164	R-PLACE1007218	16218

[0019] Furthermore, the present invention relates to the use of the above primers, as described below.

(4) A polynucleotide which can be synthesized with the primer set of (2) or (3).
(5) A polynucleotide comprising a coding region in the polynucleotide of (4).
(6) A substantially pure protein encoded by polynucleotide of (4).
(7) A partial peptide of the protein of (6).

[0020] In addition, the present invention comprises a polynucleotide described below and a protein encoded by the polynucleotide.

(8) An isolated polynucleotide selected from the group consisting of

(a) a polynucleotide comprising a coding region of the nucleotide sequence set forth in any one of the SEQ ID NOs in Tables 350 and 351;
(b) a polynucleotide comprising a nucleotide sequence encoding a protein comprising the amino acid sequence set forth in any one of the SEQ ID NOs in Tables 350 and 351;
(c) a polynucleotide comprising a nucleotide sequence encoding a protein comprising an amino acid sequence

selected from the amino acid sequences set forth in the SEQ ID NOs in Tables 350 and 351, in which one or more amino acids are substituted, deleted, inserted, and/or added, wherein said protein is functionally equivalent to the protein comprising said amino acid sequence selected from the amino acid sequences set forth in the SEQ ID NOs in Tables 350 and 351;

(d) a polynucleotide that hybridizes with a polynucleotide comprising a nucleotide sequence selected from the nucleotide sequences set forth in the SEQ ID NOs in Tables 350 and 351, and that comprises a nucleotide sequence encoding a protein functionally equivalent to the protein encoded by the nucleotide sequence selected from the nucleotide sequences set forth in the SEQ ID NOs in Tables 350 and 351;

(e) a polynucleotide comprising a nucleotide sequence encoding a partial amino acid sequence of a protein encoded by the polynucleotide of (a) to (d);

(f) a polynucleotide comprising a nucleotide sequence with at least 70% identity to the nucleotide sequence set forth in any one of the SEQ ID NOs in Tables 350 and 351.

(9). A substantially pure protein encoded by the polynucleotide of (8).

(10) An antibody against the protein or peptide of any one of (6), (7), and (9).

(11) A vector comprising the polynucleotide of (5) or (8).

(12) A transformant carrying the polynucleotide of (5) or (8), or the vector of (11).

(13) A transformant expressively carrying the polynucleotide of (5) or (8), or the vector of (11).

(14) A method for producing the protein or peptide of any one of (6), (7), and (9), comprising culturing the transformant of (13) and recovering the expression product.

(15) An oligonucleotide comprising the nucleotide sequence set forth in any one of the SEQ ID NOs in Tables 350 and 351 or the nucleotide sequence complementary to the complementary strand thereof, wherein said oligonucleotide comprises 15 nucleotides or more.

(16) Use of the oligonucleotide of (15) as a primer for synthesizing a polynucleotide.

(17) Use of the oligonucleotide of (15) as a probe for detecting a gene.

(18) An antisense polynucleotide against the polynucleotide of (8), or the portion thereof.

(19) A method for synthesizing a polynucleotide, the method comprising:

a) synthesizing a complementary strand using a cDNA library as a template, and using the primer set of (2) or (3), or the primer of (16); and

b) recovering the synthesized product.

(20) The method of (19), wherein the cDNA library is obtainable by oligo-capping method.

(21) The method of (19), wherein the complementary strand is obtainable by PCR.

(22) A method for detecting the polynucleotide of (8), the method comprising:

a) incubating a target polynucleotide with the oligonucleotide of (15) under the conditions where hybridization occurs, and

b) detecting the hybridization of the target polynucleotide with the oligonucleotide of (15).

(23) A database of polynucleotides and/or proteins, the database comprising information on at least one sequence selected from the nucleotide sequences set forth in the SEQ ID NOs in Tables 350 and 351 and/or the amino acid sequences set forth in the SEQ ID NOs in Tables 350 and 351, or a medium on which the database is stored.

[0021] Any patents, patent applications, and publications cited herein are incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0022] Figure 1 shows the restriction maps of vectors pME18SFL3 and pUC19FL3.

[0023] Figure 2 shows the reproducibility of gene expression analysis. The respective intensities of gene expression observed in independent set of experiments are plotted in the vertical axis as well as in the horizontal axis.

[0024] Figure 3 shows the detection limit in gene expression analysis. The intensity of expression is shown in the vertical axis and the concentration (μg/ml) of probe used is shown in the horizontal axis.

DETAILED DESCRIPTION OF THE INVENTION

[0025] Herein, "polynucleotide" is defined as a molecule in which multiple nucleotides are polymerized. There are no limitations in the number of the polymerized nucleotides. In case that the polymer contains relatively low number

of nucleotides, it is also described as an "oligonucleotide". The polynucleotide or the oligonucleotide of the present invention can be a natural or chemically synthesized product. Alternatively, it can be synthesized using a template DNA by an enzymatic reaction such as PCR.

[0026] All the cDNA provided by the invention are full-length cDNA. Herein, a "full-length cDNA" is defined as a cDNA which contains both ATG codon (the translation start site) and the stop codon. Accordingly, the untranslated regions, which are originally found in the upstream or downstream of the protein coding region in natural mRNA, may or may not be contained.

[0027] An "isolated polynucleotide" is a polynucleotide the structure of which is not identical to that of any naturally occurring nucleic acid or to that of any fragment of a naturally occurring genomic nucleic acid spanning more than three separate genes. The term therefore covers, for example,

(a) a DNA which has the sequence of part of a naturally occurring genomic DNA molecule but is not flanked by both of the coding sequences that flank that part of the molecule in the genome of the organism in which it naturally occurs;

(b) a nucleic acid incorporated into a vector or into the genomic DNA of a prokaryote or eukaryote in a manner such that the resulting molecule is not identical to any naturally occurring vector or genomic DNA;

(c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction fragment; and

(d) a recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein. Specifically excluded from this definition are nucleic acids present in mixtures of different (i) DNA molecules, (ii) transfected cells, or (iii) cell clones: e.g., as these occur in a DNA library such as a cDNA or genomic DNA library.

[0028] The term "substantially pure" as used herein in reference to a given polypeptide means that the protein or polypeptide is substantially free from other biological macromolecules. The substantially pure protein or polypeptide is at least 75% (e.g., at least 80, 85, 95, or 99%) pure by dry weight. Purity can be measured by any appropriate standard method, for example, by column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

[0029] All the clones (5602 clones) of the present invention are novel and encode the full-length proteins. All the clones were prepared by oligo capping method, which can achieve cDNA cloning with high fullness ratio. The cDNA clones were selected by using ATGprl score as an index of the fullness ratio at the 5'-end, based on the sequence features of the 5'-end sequences. Selection was further carried out by searching GenBank database for EST sequences homologous to 5'-end sequence of each clone by BLAST [S.F. Altschul, W. Gish, W. Miller, E.W. Myers & D.L. Lipman J. Mol. Biol., 215:403-410 (1990); W. Gish, & D.J. States, Nature Genet., 3:266-272 (1993)] and by considering the number of matching (identical) EST sequences or the number of continuous amino acids in the 5'-end sequence initiated from the initiation codon.

[0030] Moreover, the clones were turn out to be not identical to any of the known human mRNA (namely novel) by homology search using the 5'-end sequence.

[0031] The primers of the present invention, which are used for synthesizing full-length cDNA, are selected from the group comprising SEQ ID NO: 1-5547 (5'-primer), or SEQ ID NO: 5548-10463 (3'-primer). Further, the primers of the present invention, which are used for synthesizing full-length cDNA, are selected from SEQ ID NO: 16111-16164 (5'-primer), or SEQ ID NO: 16165-16218 (3'-primer). Some of the nucleotides include a known EST as its part. However, the primers of the present invention are novel in terms that the primers enable to synthesize full-length cDNA. Because the known ESTs lack important information on what part of cDNA the ESTs correspond to, it is impossible to design primers on the basis of the ESTs.

[0032] All the full-length cDNA of the present invention can be synthesized using a primer set comprising the nucleotide sequences selected from both the 5'-and 3'-end sequences, or a set comprising a primer based on the 5'-end sequence and an oligo-dT primer, by a method such as PCR (Current protocols in Molecular Biology (1987) Ausubel et al. edit, John Wiley & Sons, Section 6.1-6.4).

[0033] Specifically, PCR can be performed using an oligonucleotide that has 15 nucleotides longer, and specifically hybridizes with the complementary strand of the polynucleotide that contains the nucleotide sequence selected from the 5'-end sequences shown in Table 1 and 2 (SEQ ID NO: 1-5547, or SEQ ID NO: 16111-16164), and an oligo-dT primer as a 5'-, and 3'-primer, respectively. The length of the primers is usually 15-100 bp, and favorably between 15-35 bp. In case of LA PCR, which is described below, the primer length of 25-35 bp may provide a good result.

[0034] A method to design a primer that enables a specific amplification based on the given nucleotide sequence is known to those skilled in the art (Current Protocols in Molecular Biology, Ausubel et al. edit, (1987) John Wiley & Sons, Section 6.1-6.4). In designing a primer based on the 5'-end sequence, the primer is designed so as that, in principle, the amplification products will include the translation start site. Accordingly, in case that a given 5'-end nucleotide sequence is the 5'- untranslated region (5'UTR), any part of the sequence can be used as a 5'-primer as far as the specificity toward the target cDNA is insured. The translation start site can be predicted using a known method such

as the ATGpr as described below.

[0035] When synthesizing a polynucleotide, the target nucleotide sequence to be amplified can extend to several thousand bp in some cDNA. However, it is possible to amplify such a long nucleotides by using such as LA PCR (Long and Accurate PCR). It is advantageous to use LA PCR when synthesizing long DNA. In LA PCR, in which a special DNA polymerase having 3'→5' exonuclease activity is used, misincorporated nucleotides can be removed. Accordingly, accurate synthesis of the complementary strand can be achieved even with a long nucleotide sequence. By using LA PCR, it is reported that amplification of a nucleotide with 20 kb longer can be achieved under desirable condition (Takeshi Hayashi (1996) Jikken-Igaku Bessatsu, "Advanced Technologies in PCR" Youdo-sha).

[0036] A template DNA for synthesizing the cDNA of the present invention can be obtained by using cDNA libraries that are prepared by various methods. The full-length cDNA clones obtained here are those with high fullness ratio, which were obtained using a combination of (1) a method to prepare a full-length-enriched cDNA library using the oligo-capping method, and (2) an estimation system for fullness using the 5'-end sequence (selection based on the estimation by the ATGpr after removing clones that are not-full-length compared to the ESTs). However, it is possible to easily obtain a full-length cDNA by using the primers that are provided by the present invention, not by the above described specialized method.

The problem with the cDNA libraries prepared by the known methods or commercially available is that mRNA contained in the libraries has very low fullness ratio. Thus, it is difficult to screen full-length cDNA clone directly from the library using ordinary cloning methods. The present invention has revealed a primer that is capable of synthesizing a full-length cDNA. If provided with primers, it is possible to synthesize a target full-length cDNA by using enzymatic reactions such as PCR. In particular, a full-length-enriched cDNA library, synthesized by methods such as oligo-capping, is desirable to synthesize a full-length cDNA with more reliability.

[0037] The 5'-end sequence of the full-length cDNA clones of the invention can be used to isolate the regulatory element of transcription including the promoter on the genome. By the spring of the year 2000, a rough draft of the human genome (analysis of human genomic sequence with lower accuracy), which covers 90% of the genome, is planned to be accomplished, and by the year 2003, analysis of the entire human genomic sequence is going to be finished. However, it is hard to analyze with software the transcription start sites on the human genome, in which long introns exist. By contrast, it is easy to specify the transcription start site on the genomic sequence using the 5'-end sequence of the full-length cDNA clone, thus it is easy to obtain the genomic region involved in transcription regulation, which includes the promoter that is contained in the upstream of the transcription start site.

[0038] The full-length cDNAs cloned in the present invention are classified into 13 groups, based on the data such as ATGpr1 score, by which the fullness ratio can be evaluated. Specifically, the 13 groups consist of; the below-mentioned groups (1)-(3), containing 3690 clones (Table 9), and the group (12), containing 3 clones, wherein ATGpr1 (score defined in the ATGpr program) is higher than 0.3; and the below-mentioned groups (4)-(11), containing 1857 clones (Table 10), and the group (13), containing 52 clones, wherein, although ATGpr1 is 0.3 or less, the clones are judged to be full-length from various viewpoints. Names of the clones belonging to the groups (1)-(13) are as indicated in Examples or below.

(1) 1516 clones

Among the 3690 clones that have the maximal ATGpr1 score higher than 0.3, 1516 clones are novel full-length clones, in which at least either of the sequences of the 5'- and 3'-ends, or both are not identical to those of any human EST.

(2)377 clones

Among the 3690, 377 clones are novel full-length clones, in which the number of human EST having identical sequence at both 5'- and 3'-ends is 1 to 5.

(3) 1797 clones

Among the 3690, 1797 clones are novel full-length clones, in which the number of human EST having identical sequence at the 5'-end is not more than 20 (except the clones described above).

(4) 453 clones

Among the 1857 clones in which the maximal ATGpr1 score is 0.3 or less, the following 453 clones are estimated to be novel full-length clones since the clones have the maximal score 0.3 or more in the ATGpr2, and at least either of the sequences of their 5'- and 3'-ends, or both are not identical to those of any human EST. The ATGpr2 score is determined by using the ATGpr program with neglecting the information of the frequency of the six nucleotides contained within the sequence between the ATG codon and the stop codon (the maximal length is 300 nucleotides from the ATG codon) (Salamo A.A., Nishikawa T., and Swindells M.B. (1998) Bioinformatics, 14: 384-390; http://www.hri.co.jp/atgpr/). The ATGpr program for calculating the ATGpr2 score is described as the

ATGpr2 program in the followings.

(5) 24 clones

Among the 1857 clones, 24 clones are estimated to be full-length since their maximal ATGpr2 scores are higher than 0.3, and also novel, though they have low scores in ATGpr1 program, in which the number of the human EST having identical sequence at both 5'- and 3'-ends is 1 to 5.

(6) 65 clones

Among the 1857 clones, 65 clones are estimated to be full-length since, though they have low scores in both programs, ATGpr1 and ATGpr2, the scores are the maximum in comparison to those of the other clones in the same cluster (at least two clones). The clones are also novel, if at least either of the sequences of the 5'- and 3'-ends, or both are not identical to those of any human EST.

(7) 32 clones

Among the 1857 clones, 32 clones are estimated to be full-length since, though they have low scores in both programs, ATGpr1 and ATGpr2, the scores are the maximum in comparison to those of the other clones in the same cluster (at least two clones). The clones are also novel, if the number of the human EST having identical sequence at both 5'- and 3'-ends is 1 to 5.

(8) 36 clones

Among the 1857 clones, 36 clones are full-length, which were selected by assembling the sequences of the other clones or human EST, although they have low scores in both programs, ATGpr1 and ATGpr2. The clones are also novel, if at least either of the sequences of the 5'- and 3'-ends, or both are not identical to those of any human EST.

(9) 81 clones

Among the 1857 clones, 81 clones are full-length, which were selected by assembling the sequences of the other clones or human EST, although they have low scores in both programs, ATGpr1 and ATGpr2. The clones are also novel, if the number of the human EST having identical sequence at the 5'-end is not more than 20 (other than the clones in which at least either of the sequences of the 5'- and 3'-ends, or both are not identical to those of any human EST).

(10) 938 clones

Among the 1857 clones, 938 clones are estimated to be full-length according to the fullness ratio shown in Table 4, although they have low scores in both programs, ATGpr1 and ATGpr2. The clones are also novel, if at least the sequence of the 5'-end is not identical to those of any human EST.

(11) 228 clones

Among the 1857 clones, 228 clones are estimated to be full-length according to the fullness ratio shown in Table 7, although they have low scores in both programs, ATGpr1 and ATGpr2. The clones are also novel, if at least the sequence of the 3'-end is not identical to those of any human EST.

(12) 3 clones

Three clones, HEMBA1006812, HEMBB1001871, and NT2RP3001282, whose maximal ATGpr1 values are higher than 0.3, are full-length and novel clones whose 5'-end sequences presumably contain a coding region which is initiated with ATG codon and which encodes 100 amino acids or more.

(13) 52 clones

The following 52 clones, which have maximal ATGpr1 values of 0.3 or less, are full-length with the fullness ratios shown in Table 4 although the fullness ratios are low:

HEMBA1000497, HEMBA1001750, HEMBA1003854, HEMBA1004193, HEMBA1004860, HEMBA1005572, HEMBA1006038, HEMBA1006092, HEMBA1006406, HEMBA1006650, HEMBB1000672, HEMBB1001197, MAMMA1001252, MAMMA1002094, NT2RM4000634, NT2RM4000657, NT2RM4000783, NT2RM4000857, NT2RM4001178, NT2RM4002420, NT2RP2000198, NT2RP2000551, NT2RP2000660, NT2RP2001214, NT2RP2001460, NT2RP2001756, NT2RP2002056, NT2RP2002677, NT2RP2002755, NT2RP2002843, NT2RP2003101, NT2RP2003799, NT2RP2004095, NT2RP2004732, NT2RP2004920, NT2RP2005454, NT2RP2005776, NT2RP2005806, NT2RP2005882, NT2RP3001723, NT2RP3002099, NT2RP3003155, NT2RP3004028, OVARC1000008, OVARC1000724, OVARC1000751, OVARC1001029, PLACE1000814,

PLACE1003030, PLACE1005549, PLACE1007218, NT2RP4002298.

Moreover, the clones are novel clones whose 5' -end sequences presumably contain a coding region which is initiated with ATG codon and which encodes 50 amino acids or more. Among them, the following 20 clones is predicted to contain a coding region with 100 amino acids or more and should encode proteins:

HEMBA1000497, HEMBA1003854, HEMBA1004193, NT2RM4000657, NT2RM4001178, NT2RP2001756, NT2RP2002677, NT2RP2002755, NT2RP2002843, NT2RP2004095, NT2RP2004920, NT2RP2005806, NT2RP3002099, NT2RP3003155, OVARC1000724, OVARC1001029, PLACE1000814, PLACE1003030, PLACE1005549, PLACE1007218.

[0039] The protein encoded by the polynucleotide of the invention can be prepared as a recombinant protein or as a natural protein. For example, the recombinant protein can be prepared by inserting the polynucleotide encoding the protein of the invention into a vector, introducing the vector into an appropriate host cell and purifying the protein expressed within the transformed host cell, as described below. In contrast, the natural protein can be prepared, for example, by utilizing an affinity column to which an antibody against the protein of the invention (Current Protocols in Molecular Biology (1987) Ausubel et al. edit, John Wily & Sons, Section 16.1-16.19) is attached. The antibody used for affinity purification may be either a polyclonal antibody, or a monoclonal antibody. Alternatively, in vitro translation (See, for example, "On the fidelity of mRNA translation in the nuclease-treated rabbit reticulocyte lysate system." Dasso M.C., and Jackson R.J. (1989) Nucleic Acids Res. 17: 3129-3144) may be used for preparing the protein of the invention.

[0040] Proteins functionally equivalent to the proteins of the present invention can be prepared based on the activities, which were clarified in the above-mentioned manner, of the proteins of the present invention. Using the biological activity possessed by the protein of the invention as an index, it is possible to verify whether or not a particular protein is functionally equivalent to the protein of the invention by examining whether or not the protein has said activity.

[0041] Proteins functionally equivalent to the proteins of the present invention can be prepared by those skilled in the art, for example, by using a method for introducing mutations into an amino acid sequence of a protein (for example, site-directed mutagenesis (Current Protocols in Molecular Biology, edit, Ausubel et al., (1987) John Wiley & Sons, Section 8.1-8.5). Besides, such proteins can be generated by spontaneous mutations. The present invention comprises the proteins having one or more amino acids substitutions, deletions, insertions and/or additions in the amino acid sequences of the proteins of the present invention (Tables 350 and 351), as far as the proteins have the equivalent functions to those of the proteins identified in the present Examples described later.

[0042] There are no limitations in the number and sites of amino acid mutations, as far as the proteins maintain the functions thereof. The number of mutations is typically 30% or less, or 20% or less, or 10% or less, preferably within 5% or less, or 3% or less of the total amino acids, more preferably within 2% or less or 1 % or less of the total amino acids. From the viewpoint of maintaining the protein function, it is preferable that a substituted amino has a similar property to that of the original amino acid. For example, Ala, Val, Leu, Ile, Pro, Met, Phe and Trp are assumed to have similar properties to one another because they are all classified into a group of non-polar amino acids. Similarly, substitution can be performed among non-charged amino acid such as Gly, Ser, Thr, Cys, Tyr, Asn, and Gln, acidic amino acids such as Asp and Glu, and basic amino acids such as Lys, Arg, and His.

[0043] In addition, proteins functionally equivalent to the proteins of the present invention can be isolated by using techniques of hybridization or gene amplification known to those skilled in the art. Specifically, using the hybridization technique (Current Protocols in Molecular Biology, edit, Ausubel et al., (1987) John Wiley & Sons, Section 6.3-6.4)), those skilled in the art can usually isolate a DNA highly homologous to the DNA encoding the protein identified in the present Example based on the identified nucleotide sequence (Tables 350 and 351) or a portion thereof and obtain the functionally equivalent protein from the isolated DNA. The present invention include proteins encoded by the DNAs hybridizing with the DNAs encoding the proteins identified in the present Example, as far as the proteins are functionally equivalent to the proteins identified in the present Example. Organisms from which the functionally equivalent proteins are isolated are illustrated by vertebrates such as human, mouse, rat, rabbit, pig and bovine, but are not limited to these animals.

[0044] Washing conditions of hybridization for the isolation of DNAs encoding the functionally equivalent proteins are usually "1 × SSC, 0.1% SDS, 37°C"; more stringent conditions are "0.5 × SSC, 0.1% SDS, 42°C"; and still more stringent conditions are "0.1 × SSC, 0.1% SDS, 65°C". Alternatively, the following conditions can be given as hybridization conditions of the present invention. Namely, conditions in which the hybridization is done at "6 × SSC, 40% Formamide, 25°C", and the washing at "1 × SSC, 55°C" can be given. More preferable conditions are those in which the hybridization is done at "6 × SSC, 40% Formamide, 37°C", and the washing at "0.2 × SSC, 55°C". Even more preferable are those in which the hybridization is done at "6 × SSC, 50% Formamide, 37°C", and the washing at "0.1 × SSC, 62°C". The more stringent the conditions of hybridization are, the more frequently the DNAs highly homologous to the probe sequence are isolated. Therefore, it is preferable to conduct hybridization under stringent conditions. Examples of stringent conditions in the present invention are, washing conditions of "0.5 × SSC, 0.1% SDS, 42°C", or alternatively, hybridization conditions of "6 × SSC, 40% Formamide, 37°C", and the washing at "0.2 × SSC, 55°C". However, the above-mentioned combinations of SSC, SDS and temperature conditions are indicated just as examples.

Those skilled in the art can select the hybridization conditions with similar stringency to those mentioned above by properly combining the above-mentioned or other factors (for example, probe concentration, probe length and duration of hybridization reaction) that determines the stringency of hybridization.

[0045] The amino acid sequences of proteins isolated by using the hybridization techniques usually exhibit high homology to those of the proteins of the present invention, which are shown in Tables 350 and 351. The present invention encompasses a polynucleotide comprising a nucleotide sequence that has a high identity to the nucleotide sequence of claim 8 (a).

Furthermore, the present invention encompasses a peptide, or protein comprising an amino acid sequence that has a high identity to the amino acid sequence encoded by the polynucleotide of claim 8 (b). The term "high identity" indicates sequence identity of at least 40% or more;

preferably 60% or more; and more preferably 70% or more. Alternatively, more preferable is identity of 90% or more, or 93% or more, or 95% or more, furthermore, 97% or more, or 99% or more. The identity can be determined by using the BLAST search algorithm.

[0046] With the gene amplification technique (PCR) (Current Protocols in Molecular Biology, edit, Ausubel et al., (1987) John Wiley & Sons, Section 6.3-6.4)) using primers designed based on the nucleotide sequence (Tables 350 and 351) or a portion thereof identified in the present Example, it is possible to isolate a DNA fragment highly homologous to the polynucleotide sequence or a portion thereof and to obtain functionally equivalent protein to a particular protein identified in the present Example based on the isolated DNA fragment.

[0047] The "percent identity" of two amino acid sequences or of two nucleic acids is determined using the algorithm of Karlin and Altschul (Proc. Natl. Acad. Sei. USA 87:2264-2268, 1990), modified as in Karlin and Altschul (Proc. Natl. Acad. Sei. USA 90:5873-5877, 1993). Such an algorithm is incorporated into the BLASTN and BLASTX programs of Altschul et al. (J. Mol. Biol.215:403-410, 1990). BLAST nucleotide searches are performed with the BLASTN program, score = 100, wordlength = 12. BLAST protein searches are performed with the BLASTX program, score = 50, wordlength = 3. When gaps exist between two sequences, Gapped BLAST is utilized as described in Altschul et al. (Nucleic Acids Res.25:3389-3402,1997). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., BLASTX and BLASTN) are used. See http://www.ncbi.nlm.nih.gov.

[0048] The present invention also includes a partial peptide of the proteins of the invention. The partial peptide comprises a protein generated as a result that a signal peptide has been removed from a secretory protein. If the protein of the present invention has an activity as a receptor or a ligand, the partial peptide may function as a competitive inhibitor of the protein and may bind to the receptor (or ligand). In addition, the present invention comprises an antigen peptide for raising antibodies. For the peptides to be specific for the protein of the invention, the peptides comprise at least 7 amino acids, preferably 8 amino acids or more, more preferably 9 amino acids or more, and even more preferably 10 amino acids or more. The peptide can be used for preparing antibodies against the protein of the invention, or competitive inhibitors of them, and also screening for a receptor that binds to the protein of the invention. The partial peptides of the invention can be produced, for example, by genetic engineering methods, known methods for synthesizing peptides, or digesting the protein of the invention with an appropriate peptidase.

[0049] The present invention also relates to a vector into which the DNA of the invention is inserted. The vector of the invention is not limited as long as it contains the inserted DNA stably. For example, if *E. coli* is used as a host, vectors such as pBluescript vector (Stratagene) are preferable as a cloning vector. To produce the protein of the invention, expression vectors are especially useful. Any expression vector can be used as far as it is capable of expressing the protein *in vitro*, in *E. coli*, in cultured cells, or *in vivo*. For example, pBEST vector (Promega) is preferable for *in vitro* expression, pET vector (Invitrogen) for *E. coli*, pME18S-FL3 vector (GenBank Accession No. AB009864) for cultured cells, and pME18S vector (Mol. Cell. Biol. (1988) 8: 466-472) for *in vivo* expression. To insert the DNA of the invention, ligation utilizing restriction sites can be performed according to the standard method (Current Protocols in Molecular Biology (1987) Ausubel et al. edit, John Wily & Sons, Section 11.4-11.11).

[0050] The present invention also relates to a transformant carrying the vector of the invention. Any cell can be used as a host into which the vector of the invention is inserted, and various kinds of host cells can be used depending on the purposes. For strong expression of the protein in eukaryotic cells, COS cells or CHO cells can be used, for example.

[0051] Introduction of the vector into host cells can be performed, for example, by calcium phosphate precipitation method, electroporation method (Current Protocols in Molecular Biology (1987) Ausubel et al. edit, John Wily & Sons, Section 9.1-9.9), lipofectamine method (GIBCO-BRL), or microinjection method, etc.

[0052] The primer of the present invention can be used for synthesizing full-length cDNA, and also for the detection and/or diagnosis of the abnormality of the protein of the invention encoded by the full-length cDNA. For example, by utilizing polymerase chain reaction (genomic DNA-PCR, or RT-PCR) using the primer of the invention, DNA encoding the protein of the invention can be amplified. It is also possible to obtain the regulatory region of expression in the 5'-upstream by using PCR or hybridization since the transcription start site within the genomic sequence can be easily specified based on the 5'-end sequence of the full-length cDNA. The obtained genomic region can be used for detection and/or diagnosis of the abnormality of the sequence by RFLP analysis, SSCP, or direct sequencing.

[0053] Furthermore, the "polynucleotide having a length of at least 15 nucleotides, comprising a nucleotide sequence that is complementary to a polynucleotide comprising the nucleotide sequence set forth in any one of SEQ ID NOs in Tables 350 and 351, or its complementary strand" includes an antisense polynucleotide for suppressing the expression of the protein of the invention. To exert the antisense effect, the antisense polynucleotide has a length of at least 15 bp or more, for example, 50 bp or more, preferably 100 bp or more, and more preferably 500 bp or more, and has a length of usually 3000 bp or less and preferably 2000 bp or less. The antisense DNA can be used in the gene therapy of the diseases that are caused by the abnormality of the protein of the invention (abnormal function or abnormal expression). Said antisense DNA can be prepared, for example, by the phosphorothioate method ("Physicochemical properties of phosphorothioate oligodeoxynucleotides." Stein (1988) Nucleic Acids Res. 16: 3209-3221) based on the nucleotide sequence of the DNA encoding the protein (for example, the DNA set forth in any one of SEQ ID NOs in Tables 350 and 351).

[0054] The polynucleotide or antisense DNA of the present invention can be used in gene therapy, for example, by administrating it into a patient by the in vivo or ex vivo method with virus vectors such as retrovirus vectors, adenovirus vectors, and adeno-associated virus vectors, or non-virus vectors such as liposome.

[0055] The present invention also relates to antibodies that bind to the protein of the invention. There are no limitations in the form of the antibodies of the invention. They include polyclonal antibodies, monoclonal antibodies, or their portions that can bind to the protein of the invention. They also include antibodies of all classes. Furthermore, special antibodies such as humanized antibodies are also included.

[0056] The polyclonal antibody of the invention can be obtained according to the standard method by synthesizing an oligopeptide corresponding to the amino acid sequence and immunizing rabbits with the peptides (Current Protocols in Molecular Biology (1987) Ausubel et al. edit, John Wily & Sons, Section 11.12-11.13). The monoclonal antibody of the invention can be obtained according to the standard method by purifying the protein expressed in *E. coli*, immunizing mice with the protein, and producing a hybridoma cell by fusing the spleen cells and myeloma cells (Current Protocols in Molecular Biology (1987) Ausubel et al. edit, John Wily & Sons, Section 11.4-11.11).

[0057] The antibody binding to the protein of the present invention can be used for purification of the protein of the invention, and also for detection and/or diagnosis of the abnormalities of the expression and structure of the protein. Specifically, proteins can be extracted, for example, from tissues, blood, or cells, and the protein of the invention is detected by Western blotting, immunoprecipitation, or ELISA, etc. for the above purpose.

[0058] Furthermore, the antibody binding to the protein of the present invention can be utilized for treating the diseases that associates with the protein of the invention. If the antibodies are used for treating patients, human antibodies or humanized antibodies are preferable in terms of their low antigenicity. The human antibodies can be prepared by immunizing a mouse whose immune system is replaced with that of human ("Functional transplant of megabase human immunoglobulin loci recapitulates human antibody response in mice" Mendez M.J. et al. (1997) Nat. Genet. 15: 146-156). The humanized antibodies can be prepared by recombination of the hypervariable region of a monoclonal antibody (Methods in Enzymology (1991) 203: 99-121).

[0059] The cDNA of the present invention encodes the amino acid sequence of a protein which is predicted to have the function(s) described below based on the homology search of the GenBank and SwissProt. Specifically, for instance, as shown in EXAMPLES, searching a known gene or protein that is homologous to the partial sequence of the full-length cDNA of the invention (5602 clone) and referring the function of the gene and of the protein encoded by the gene make it possible to predict the function of the protein encoded by the cDNA of the invention. In this way, each of 1437 clones out of the 5602 full-length cDNA clones of the invention was predicted to encode a protein that was classified into one or more of the following categories.

Secretory or membrane protein (261 clones)
Glycoprotein-associated protein (113 clones)
Signal transduction-associated protein (148 clones)
Transcription-associated protein (233 clones)
Disease-associated protein (437 clones)
Enzyme or metabolism-associated protein (301 clones)
Cell division- or cell proliferation-associated protein (74 clones)
Cytoskeleton-associated protein (92 clones)
RNA synthesis-associated protein (280 clones)
Nuclear protein (352 clones)
Protein synthesis- or transport-associated protein (112 clones)
Cellular defense-associated protein (23 clones)
Development- or growth-associated protein (23 clones)

[0060] It is also possible to predict the protein function by looking into the amino acid sequence for the motifs such

as the signal sequence, transmembrane region, nuclear translocation signal, glycosylation signal, phosphorylation site, Zinc finger motif, and SH3 domain. The programs, PSORT (Nakai K., and Kanehisa M. (1992) Genomics 14: 897-911), SOSUI (Hirokawa T. et al. (1998) Bioinformatics 14. 378-379) (Mitsui Information Developing Inc.), and MEMSAT (Jones D.T., Taylor W.R., and Thornton J.M. (1994) Biochemistry 33: 3038-3049) can be used to predict the existence of the signal sequence or transmembrane region. Alternatively, a partial amino acid sequence of the protein is fused with another protein such as GFP, the fusion protein is transfected into cultured cells, and the localization is analyzed to predict the function of the original protein.

[0061] Based on the determined nucleotide sequences of the full-length cDNAs obtained in the present invention, it is possible to predict more detailed functions of the proteins encoded by the cDNA clones, for example, by searching the databases such as GenBank, Swiss-Prot and UniGene for homologies of the cDNAs; or by searching the amino acid sequences deduced from the full-length cDNAs for signal sequences by using software programs such as PSORT, for transmembrane regions by using software programs such as SOSUI or for motifs by using software programs such as Pfam (http://www.sanger.ac.uk/Software/Pfam/index.shtml) and PROSITE (http://www.expasy.ch/prosite/). As a matter of course, the functions are often predictable by using partial sequence information (preferably 300 nucleotides or more) instead of the full-length nucleotide sequences. However, the result of the prediction by using partial nucleotide sequence does not always agree with the result obtained by using full-length nucleotide sequence, and thus, it is needless to say that the prediction of function is preferably performed based on the full-length nucleotide sequences. GenBank, Swiss-Prot and UniGene databases were searched for homologies of the full-length nucleotide sequences of the 4997 clones (see Example 18). The amino acid sequences deduced from the full-length nucleotide sequences were searched for functional domains by PSORT, SOSUI and Pfam. Prediction of functions of proteins encoded by the clones and the categorization thereof were performed based on these results obtained.

The following 798 clones were categorized into secretory and/or membrane proteins.

HEMBA1000356, HEMBA1000518, HEMBA1000531, HEMBA1000637, HEMBA1000719, HEMBA1000817,
HEMBA1000822, HEMBA1000852, HEMBA1000870, HEMBA1000991, HEMBA1001052, HEMBA1001071,
HEMBA1001085, HEMBA1001286, HEMBA1001351, HEMBA1001407, HEMBA1001446, HEMBA1001515,
HEMBA1001557, HEMBA1001569, HEMBA1001661, HEMBA1001734, HEMBA1001746, HEMBA1001866,
HEMBA1002125, HEMBA1002150, HEMBA1002166, HEMBA1002417, HEMBA1002462, HEMBA1002475,
HEMBA1002477, HEMBA1002486, HEMBA1002609, HEMBA1002659, HEMBA1002661, HEMBA1002780,
HEMBA1002818, HEMBA1002876, HEMBA1002921, HEMBA1003071, HEMBA1003077, HEMBA1003079,
HEMBA1003086, HEMBA1003096, HEMBA1003281, HEMBA1003286, HEMBA1003538, HEMBA1003711,
HEMBA1003742, HEMBA1003803, HEMBA1004055, HEMBA1004143, HEMBA1004146, HEMBA1004207,
HEMBA1004341, HEMBA1004461, HEMBA1004577, HEMBA1004637, HEMBA1004752, HEMBA1004756,
HEMBA1004850, HEMBA1004889, HEMBA1004923, HEMBA1004930, HEMBA1005029, HEMBA1005035,
HEMBA1005050, HEMBA1005552, HEMBA1005576, HEMBA1005581, HEMBA1005588, HEMBA1005616,
HEMBA1005699, HEMBA1005991, HEMBA1006036, HEMBA1006038, HEMBA1006067, HEMBA1006173,
HEMBA1006198, HEMBA1006293, HEMBA1006310, HEMBA1006492, HEMBA1006502, HEMBA1006583,
HEMBA1006659, HEMBA1006758, HEMBA1006789, HEMBA1006921, HEMBA1006926, HEMBA1006976,
HEMBA1007203, HEMBA1007301, HEMBB1000037, HEMBB1000050, HEMBB1000054, HEMBB1000175,
HEMBB1000317, HEMBB1000556, HEMBB1000593, HEMBB1000631, HEMBB1000763, HEMBB1000827,
HEMBB1000915, HEMBB1000975, HEMBB1001112, HEMBB1001151, HEMBB1001177, HEMBB1001302,
HEMBB1001348, HEMBB1001564, HEMBB1001630, HEMBB1001871, HEMBB1001872, HEMBB1001925,
HEMBB1001962, HEMBB1002042, HEMBB1002044, HEMBB1002142, HEMBB1002190, HEMBB1002193,
HEMBB1002247, HEMBB1002383, HEMBB1002387, HEMBB1002550, HEMBB1002600, HEMBB1002692,
MAMMA1000045, MAMMA1000129, MAMMA1000133, MAMMA1000277, MAMMA1000278, MAMMA1000410,
MAMMA1000416, MAMMA1000472, MAMMA1000672, MAMMA1000684, MAMMA1000714, MAMMA1000734,
MAMMA1000778, MAMMA1000798, MAMMA1000842, MAMMA1000859, MAMMA1000897, MAMMA1000956,
MAMMA1001008, MAMMA1001030, MAMMA1001041, MAMMA1001073, MAMMA1001080, MAMMA1001139,
MAMMA1001154, MAMMA1001322, MAMMA1001388, MAMMA1001411, MAMMA1001487, MAMMA1001751,
MAMMA1001754, MAMMA1001771, MAMMA1002009, MAMMA1002427, MAMMA1002428, MAMMA1002461,
MAMMA1002524, MAMMA1002573, MAMMA1002598, MAMMA1002655, MAMMA1002684, MAMMA1002769,
MAMMA1002844, MAMMA1002881, MAMMA1002890, MAMMA1002938, MAMMA1002947, MAMMA1003035,
MAMMA1003089, MAMMA1003146, MAMMA1003150, NT2RM1000035, NT2RM1000037, NT2RM1000062,
NT2RM1000080, NT2RM1000092, NT2RM1000131, NT2RM1000199, NT2RM1000257, NT2RM1000260,
NT2RM1000355, NT2RM1000430, NT2RM1000563, NT2RM1000648, NT2RM1000742, NT2RM1000770,
NT2RM1000800, NT2RM1000811, NT2RM1000833, NT2RM1000857, NT2RM1000867, NT2RM1000882,
NT2RM1000905, NT2RM1001008,
NT2RM1001115, NT2RM1001139, NT2RM2000259, NT2RM2000260, NT2RM2000287, NT2RM2000395,
NT2RM2000402, NT2RM2000407, NT2RM2000422, NT2RM2000490, NT2RM2000522, NT2RM2000566,

PLACE1005884, PLACE1005934, PLACE1006076, PLACE1006119, PLACE1006159, PLACE1006164, PLACE1006170, PLACE1006382, PLACE1006492, PLACE1006629, PLACE1006704, PLACE1006731, PLACE1006760, PLACE1006779, PLACE1006795, PLACE1006805, PLACE1006962, PLACE1007045, PLACE1007111, PLACE1007282, PLACE1007386, PLACE1007416, PLACE1007484, PLACE1007544, PLACE1007645, PLACE1007743, PLACE1007746, PLACE1007807, PLACE1007858, PLACE1008002, PLACE1008181, PLACE1008273, PLACE1008368, PLACE1008405, PLACE1008532, PLACE1008568, PLACE1008625, PLACE1008696, PLACE1008867, PLACE1009027, PLACE1009039, PLACE1009045, PLACE1009110, PLACE1009298, PLACE1009328, PLACE1009581, PLACE1009621, PLACE1009622, PLACE1009637, PLACE1009925, PLACE1009935, PLACE1010089, PLACE1010106, PLACE1010152, PLACE1010274, PLACE1010491, PLACE1010629, PLACE1010630, PLACE1010714, PLACE1010739, PLACE1010891, PLACE1010896, PLACE1010925, PLACE1010965, PLACE1011026, PLACE1011046, PLACE1011214, PLACE1011399, PLACE1011433, PLACE1011492, PLACE1011641, PLACE1011649, PLACE1011719, PLACE1011762, PLACE1011858, PLACE1011923, PLACE2000014, PLACE2000039, PLACE2000216, PLACE2000302, PLACE2000317, PLACE2000342, PLACE2000347, PLACE2000379, PLACE3000121, PLACE3000124, PLACE3000160, PLACE3000242, PLACE3000271, PLACE3000353, PLACE3000362, PLACE3000365, PLACE3000400, PLACE3000401, PLACE4000034, PLACE4000089, PLACE4000522, PLACE4000558, SKNMC1000050, THYRO1000040, THYRO1000197, THYRO1000241, THYRO1000327, THYRO1000394, THYRO1000488, THYRO1000501, THYRO1000585, THYRO1000596, THYRO1000625, THYRO1000805, THYRO1000934, THYRO1001133, THYRO1001134, THYRO1001173, THYRO1001213, THYRO1001262, THYRO1001290, THYRO1001721, Y79AA1000037, Y79AA1000800, Y79AA1000976, Y79AA1001078, Y79AA1001228, Y79AA1001299, Y79AA1001402, Y79AA1001585, Y79AA1001696, Y79AA1001711, Y79AA1001827, Y79AA1001875, Y79AA1002027, Y79AA1002211, Y79AA1002234, Y79AA1002258

[0099] On the other hand, clones of which expression levels decrease by RA/inhibitor are as follows:

HEMBA1000012, HEMBA1000501, HEMBA1000946, HEMBA1003220, HEMBA1003403, HEMBA1003569, HEMBA1003591, HEMBA1003926, HEMBA1004168, HEMBA1004507, HEMBA1005009, HEMBA1005296, HEMBA1005528, HEMBA1005570, HEMBA1006467, HEMBA1006486, HEMBA1006492, HEMBA1007322, HEMBB1000055, HEMBB1000244, HEMBB1001665, MAMMA1000684, MAMMA1001139, MAMMA1001743, NT2RM1000257 NT2RM1000318, NT2RM1000539, NT2RM1000666, NT2RM2000092, NT2RM2000192, NT2RM2000371, NT2RM2000594, NT2RM4000511, NT2RM4001140, NT2RM4001754, NT2RM4001905, NT2RM4001940, NT2RM4002593, NT2RP1000086, NT2RP1000439, NT2RP1001073, NT2RP2000098, NT2RP2000965, NT2RP2001397, NT2RP2002047, NT2RP2004226, NT2RP2004396, NT2RP2004655, NT2RP2005126, NT2RP2005464, NT2RP2005712, NT2RP2005859, NT2RP2005890, NT2RP3000980, NT2RP3001383, NT2RP3001621, NT2RP3002081, NT2RP3002181, NT2RP3002244, NT2RP3002590, NT2RP3003059, NT2RP3004258, NT2RP3004378, NT2RP3004527, NT2RP3004594, NT2RP4001760, NT2RP4001950, NT2RP4002047, NT2RP4002408, NT2RP5003459, OVARC1000004, OVARC1000035, OVARC1000431, OVARC1001051, OVARC1001129, OVARC1001176, OVARC1001261, OVARC1001342, OVARC1001942, OVARC1001943, PLACE1002171, PLACE1002465, PLACE1003190, PLACE1003375, PLACE1004128, PLACE1005026, PLACE1005876, PLACE1005923, PLACE1007257, PLACE1007375, PLACE1007507, PLACE1008941, PLACE1010624, PLACE1011090, PLACE1011219, THYRO1000270, Y79AA1000346, Y79AA1001541

[0100] These clones are also associated with neural differentiation and, therefore, are candidates for genes associated with neurological diseases.

[0101] For example, if the protein encoded by the cDNA of the present invention is a regulatory factor of cellular conditions such as growth and differentiation, it can be used for developing medicines as follows. The protein or antibody provided by the invention is injected into a certain kind of cells by microinjection. Then, using the cells, it is possible to screen low molecular weight compounds by measuring the change in the cellular conditions, or the activation or inhibition of a particular gene. The screening can be performed as follows. First, the protein is expressed and purified as recombinant. The purified protein is microinjected into cells such as various cell lines, or primary culture cells, and the cellular change such as growth and differentiation can be examined. Alternatively, the induction of genes whose expression is known to be associated with a particular change of cellular conditions may be detected by the amount of mRNA or protein. Or, the amount of intracellular molecules (low molecular weight compounds, etc.) that is changed by the function of the gene product (protein) which is known to be associated with a particular change of cellular conditions may be detected. The compounds to be screened (both low and high molecular compounds are acceptable) can be added to the culture media and assessed for their activity by measuring the change of the cellular conditions. Instead of microinjection, cell lines introduced with the gene obtained in the invention can be used for the screening. If the gene product is turn out to be associated with a particular change in the cellular conditions, the change of the product can be used as a measurement for screening. Once a compound is screened out which can activate or inhibit

the function of the protein of the invention, it can be applied for developing medicines.

[0102] If the protein encoded by the cDNA of the present invention is a secretory protein, membrane protein, or protein associated with signal transduction, glycoprotein, transcription, or diseases, it can be used in functional assays for developing medicines.

[0103] In case of a membrane protein, it is most likely to be a protein that functions as a receptor or ligand on the cell surface. Therefore, it is possible to reveal a new relationship between a ligand and receptor by screening the membrane protein of the invention based on the binding activity with the known ligand or receptor. Screening can be performed according to the known methods.

[0104] For example, a ligand against the protein of the invention can be screened in the following manner. Namely, a ligand that binds to a specific protein can be screened by a method comprising the steps of: (a) contacting a test sample with the protein of the invention or a partial peptide thereof, or cells expressing these, and (b) selecting a test sample that binds to said protein, said partial peptide, or said cells.

[0105] On the other hand, for example, screening using cells expressing the protein of the present invention that is a receptor protein can also be performed as follows. It is possible to screen receptors that is capable of binding to a specific protein by using procedures (a) attaching the sample cells to the protein of the invention or its partial peptide, and (b) selecting cells that can bind to the said protein or its partial peptide.

[0106] In a following screening as an example, first the protein of the invention is expressed, and the recombinant protein is purified. Next, the purified protein is labeled, binding assay is performed using a various cell lines or primary cultured cells, and cells that are expressing a receptor are selected (Growth and differentiation factors and their receptors, Shin-Seikagaku Jikken Kouza Vol.7 (1991) Honjyo, Arai, Taniguchi, and Muramatsu edit, p203-236, Tokyo-Kagaku-Doujin). A protein of the invention can be labeled with RI such as ^{125}I, and enzyme (alkaline phosphatase etc.). Alternatively, a protein of the invention may be used without labeling and then detected by using a labeled antibody against the protein. The cells that are selected by the above screening methods, which express a receptor of the protein of the invention, can be used for the further screening of an agonists or antagonists of the said receptor.

[0107] Once the ligand binding to the protein of the invention, the receptor of the protein of the invention or the cells expressing the receptor are obtained by screening, it is possible to screen a compound that binds to the ligand and receptor. Also it is possible to screen a compound that can inhibit both bindings (agonists or antagonists of the receptor, for example) by utilizing the binding activities.

[0108] When the protein of the invention is a receptor, the screening method comprises the steps of (a) contacting the protein of the invention or cells expressing the protein of the invention with the ligand, in the presence of a test sample, (b) detecting the binding activity between said protein or cells expressing said protein and the ligand, and (c) selecting a compound that reduces said binding activity when compared to the activity in the absence of the test sample. Furthermore, when the protein of the invention is a ligand, the screening method comprises the steps of (a) contacting the protein of the invention with its receptor or cells expressing the receptor in the presence of samples, (b) detecting the binding activity between the protein and its receptor or the cells expressing the receptor, and (c) selecting a compound that can potentially reduce the binding activity compared to the activity in the absence of the sample.

[0109] Samples to screen include cell extracts, expressed products from a gene library, synthesized low molecular compound, synthesized peptide, and natural compounds, for example, but are not construed to be listed here. A compound that is isolated by the above screening using a binding activity of the protein of the invention can also be used as a sample.

[0110] A compound isolated by the screening may be a candidate to be an agonist or an antagonist of the receptor of the protein. By utilizing an assay that monitors a change in the intracellular signaling such as phosphorylation which results from reduction of the binding between the protein and its receptor, it is possible to identify whether the obtained compound is an agonist or antagonist of the receptor. Also, the compound may be a candidate of a molecule that can inhibit the interaction between the protein and its associated proteins (including a receptor) in vivo. Such compounds can be used for developing drugs for precaution or cures of a disease with which the protein is associated.

[0111] Secretory proteins may regulate cellular conditions such as growth and differentiation. It is possible to find out a novel factor that regulates cellular conditions by adding the secretory protein of the invention to a certain kind of cell, and performing a screening by utilizing the cellular changes in growth or differentiation, or activation of a particular gene.

[0112] The screening can be performed, for example, as follows. First, the protein of the invention is expressed and purified in a recombinant form. Then, the purified protein is added to a various kind of cell lines or primary cultured cells, and the change in the cell growth and differentiation is monitored. The induction of a particular gene that is known to be involved in a certain cellular change is detected by the amounts of mRNA and protein. Alternatively, the amount of an intracellular molecule (low-molecular-weight compounds, etc.) that is changed by the function of a gene product (protein) that is known to function in a certain cellular change is used for the detection.

Once the screening reveals that the protein of the invention can regulate cellular conditions or the functions, ...ply the protein as a pharmaceutical and diagnostic medicine for associated diseases by itself or by

altering a part of it into an appropriate composition.

[0114] As is above described for membrane proteins, the secretory protein provided by the invention may be used to explore a novel ligand-receptor interaction using a screening based on the binding activity to a known ligand or receptor. A similar method can be used to identify an agonist or antagonist. The resulting compounds obtained by the methods can be a candidate of a compound that can inhibit the interaction between the protein of the invention and an interacting molecule (including a receptor). The compounds may be able to use as a preventive, therapeutic, and diagnostic medicine for the diseases, in which the protein may play a certain role.

[0115] Proteins associated with signal transduction or transcription may be a factor that affects a certain protein or gene in response to intracellular/extracellular stimuli. It is possible to find out a novel factor that can affect a protein or gene by expressing the protein provided by the invention in a certain types of cells, and performing a screening utilizing the activation of a certain intracellular protein or gene.

[0116] The screening may be performed as follows. First, a transformed cell line expressing the protein is obtained. Then, the transformed cell line and the untransformed original cell line are compared for the changes in the expression of a certain gene by detecting the amount of its mRNA or protein. Alternatively, the amount of an intracellular molecule (low molecular weight compounds) that is changed by the function of a certain gene product (protein) may be used for the detection. Furthermore, the change of the expression of a certain gene can be detected by introducing a fusion gene that comprises a regulatory region of the gene and a marker gene (luciferase, beta-galactosidase, etc.) into a cell, expressing the protein provided by the invention into the cell, and estimating the activity of a marker gene product (protein).

[0117] If the protein or gene of the invention is associated with diseases, it is possible to screen a gene or compound that can regulate its expression and/or activity either directly or indirectly by utilizing the protein of the present invention.

[0118] For example, the protein of the invention is expressed and purified as a recombinant protein. Then, the protein or gene that interacts with the protein of the invention is purified, and screened based on the binding. Alternatively, the screening can be performed by adding with a compound of a candidate of the inhibitor added in advance and monitoring the change of binding activity. In another method, a transcription regulatory region locating in the 5'-upstream of the gene encoding the protein of the invention that is capable of regulating the expression of other genes is obtained, and fused with a marker gene. The fusion is introduced into a cell, and the cell is added with compounds to explore a regulatory factor of the expression of the said gene.

[0119] The compound obtained by the screening can be used for developing pharmaceutical and diagnostic medicines for the diseases with which the protein of the present invention is associated. Similarly, if the regulatory factor obtained in the screening is turn out to be a protein, compounds that can newly affect the expression or activity of the protein may be used as a medicine for the diseases with which the protein of the invention is associated.

[0120] If the protein of the invention has an enzymatic activity, regardless as to whether it is a secretory protein, membrane protein, or proteins associated with signal transduction, glycoprotein, transcription, or diseases, a screening may be performed by adding a compound to the protein of the invention and monitoring the change of the compound. The enzymatic activity may also be utilized to screen a compound that can inhibit the activity of the protein.

[0121] In a screening given as an example, the protein of the invention is expressed and the recombinant protein is purified. Then, compounds are contacted with the purified protein, and the amount of the compound and the reaction products is examined. Alternatively, compounds that are candidates of an inhibitor are pretreated, then a compound (substrate) that can react with the purified protein is added, and the amount of the substrate and the reaction products is examined.

[0122] The compounds obtained in the screening may be used as a medicine for diseases with which the protein of the invention is associated. Also they can be applied for tests that examine whether the protein of the invention functions normally *in vivo*.

[0123] Whether the secretory protein, membrane protein, signal transduction-associated protein, glycoprotein-associated protein, or transcription-associated protein of the present invention is a novel protein associated with diseases or not is determined in another method than described above, by obtaining a specific antibody against the protein of the invention, and examining the relationship between the expression or activity of the protein and a certain disease. In an alternative way, it may be analyzed referred to the methods in "Molecular Diagnosis of Genetic Diseases" (Elles R. edit, (1996) in the series of "Method in Molecular Biology" (Humana Press).

[0124] Proteins associated with diseases are targets of screening as mentioned, and thus are very useful in developing drugs which regulate their expression and activity. Also, the proteins are useful in the medicinal industry as a diagnostic marker of the associated disease or a target of gene therapy.

[0125] Compounds isolated as mentioned above can be administered patients as it is, or after formulated into a pharmaceutical composition according to the known methods. For example, a pharmaceutically acceptable carrier or vehicle, specifically sterilized water, saline, plant oil, emulsifier, or suspending agent can be mixed with the compounds appropriately. The pharmaceutical compositions can be administered to patients by a method known to those skilled in the art, such as intraarterial, intravenous, or subcutaneous injections. The dosage may vary depending on the weight

or age of a patient, or the method of administration, but those skilled in the art can choose an appropriate dosage properly. If the compound is encoded by DNA, the DNA can be cloned into a vector for gene therapy, and used for gene therapy. The dosage of the DNA and the method of its administration may vary depending on the weight or age of a patient, or the symptoms, but those skilled in the art can choose properly.

[0126] The present invention further relates to databases comprising at least a sequence of polynucleotide and/or protein, or a medium recorded in such databases, selected from the sequence data of the nucleotide and/or the amino acids indicated in Table 350 and Table 351.

The term "database" means a set of accumulated information as machine-searchable and readable information of nucleotide sequence. The databases of the present invention comprise at least one of the novel nucleotide sequences of polynucleotides provided by the present invention. The databases of the present invention can consist of only the sequence data of the novel polynucleotides provided by the present invention or can comprise other information on nucleotide sequences of known full-length cDNAs or ESTs. The databases of the present invention can be comprised of not only the information on the nucleotide sequences but also the information on the gene functions revealed by the present invention. Additional information such as names of DNA clones carrying the full-length cDNAs can be recorded or linked together with the sequence data in the databases.

[0127] The database of the present invention is useful for gaining complete gene sequence information from partial sequence information of a gene of interest. The database of the present invention comprises nucleotide sequence information of full-length cDNAs. Consequently, by comparing the information in this database with the nucleotide sequence of a partial gene fragment yielded by differential display method or subtraction method, the information on the full-length nucleotide sequence of interest can be gained from the sequence of the partial fragment as a starting clue.

[0128] The sequence information of the full-length cDNAs constituting the database of the present invention contains not only the information on the complete sequences but also extra information on expression frequency of the genes as well as homology of the genes to known genes and known proteins. Thus the extra information facilitates rapid functional analyses of partial gene fragments. Further, the information on human genes is accumulated in the database of the present invention, and therefore, the database is useful for isolating a human homologue of a gene originating from other species. The human homologue can be isolated based on the nucleotide sequence of the gene from the original species.

[0129] At present, information on a wide variety of gene fragments can be obtained by differential display method and subtraction method. In general, these gene fragments are utilized as tools for isolating the full-length sequences thereof. When the gene fragment corresponds to an already-known gene, the full-length sequence is easily obtained by comparing the partial sequence with the information in known databases. However, when there exists no information corresponding to the partial sequence of interest in the known databases, cDNA cloning should be carried out for the full-length cDNA. It is often difficult to obtain the full-length nucleotide sequence using the partial sequence information as an initial clue. If the full-length of the gene is not available, the amino acid sequence of the protein encoded by the gene remains unidentified. Thus the database of the present invention can contribute to the identification of full-length cDNAs corresponding to gene fragments, which cannot be revealed by using databases of known genes.

[0130] The present invention has provided 5602 novel full-length cDNA clones, and primers for synthesizing the cDNA. As has not yet proceeded the isolation of full-length cDNA within the human, the invention has great significance. The full-length cDNA clones contain the translation initiation site, and thus provide a useful information for analysis of protein functions.

[0131] The cDNA clones are assumed to encode proteins such as secretory proteins, membrane proteins, signal transduction-associated protein, glycoprotein-associated protein, or transcription-associated protein, etc., which have important functions in vivo, and also predicted to be associated with many diseases. The genes and proteins associated with diseases are useful for developing a diagnostic marker or medicines for regulation of their expression and activity, or as a target of gene therapy.

[0132] The invention is illustrated more specifically with reference to the following examples, but is not to be construed as being limited thereto.

EXAMPLE 1

Construction of a cDNA library by the oligo-capping method.

[0133] The NT-2 neuron progenitor cells (Stratagene), a teratocarcinoma cell line from human embryo testis, which can differentiate into neurons by the treatment with retinoic acid were used.

The NT-2 cells were cultured according to the manufacturer's instructions as follows.

(1) NT-2 cells were cultured without induction by retinoic acid treatment (NT2RM1, NT2RM2, NT2RM4).

(2) After cultured, NT-2 cells were induced by adding retinoic acid, and then were cultured for 48 hours (NT2RP1).

NO: 16216, SEQ ID NO: 16163 / SEQ ID NO: 16217, and SEQ ID NO: 16164 / SEQ ID NO: 16218

4. A polynucleotide which can be synthesized with the primer set of claim 2 or 3.

5. A polynucleotide comprising a coding region in the polynucleotide of claim 4.

6. A substantially pure protein encoded by polynucleotide of claim 4.

7. A partial peptide of the protein of claim 6.

8. An isolated polynucleotide selected from the group consisting of

(a) a polynucleotide comprising a coding region of the nucleotide sequence set forth in any one of the following SEQ ID NOs:

SEQ ID NO: 10468, SEQ ID NO: 10470, SEQ ID NO: 10471, SEQ ID NO: 10472, SEQ ID NO: 10473, SEQ ID NO: 10475, SEQ ID NO: 10477, SEQ ID NO: 10479, SEQ ID NO: 10481, SEQ ID NO: 10483, SEQ ID NO: 10485, SEQ ID NO: 10487, SEQ ID NO: 10488, SEQ ID NO: 10489, SEQ ID NO: 10491, SEQ ID NO: 10493, SEQ ID NO: 10495, SEQ ID NO: 10496, SEQ ID NO: 10497, SEQ ID NO: 10498, SEQ ID NO: 10500, SEQ ID NO: 10502, SEQ ID NO: 10503, SEQ ID NO: 10504, SEQ ID NO: 10505, SEQ ID NO: 10507, SEQ ID NO: 10508, SEQ ID NO: 10510, SEQ ID NO: 10511, SEQ ID NO: 10512, SEQ ID NO: 10514, SEQ ID NO: 10516, SEQ ID NO: 10517, SEQ ID NO: 10519, SEQ ID NO: 10521, SEQ ID NO: 10523, SEQ ID NO: 10524, SEQ ID NO: 10526, SEQ ID NO: 10528, SEQ ID NO: 10529, SEQ ID NO: 10530, SEQ ID NO: 10532, SEQ ID NO: 10534, SEQ ID NO: 10535, SEQ ID NO: 10537, SEQ ID NO: 10539, SEQ ID NO: 10540, SEQ ID NO: 10542, SEQ ID NO: 10543, SEQ ID NO: 10545, SEQ ID NO: 10546, SEQ ID NO: 10548, SEQ ID NO: 10550, SEQ ID NO: 10551, SEQ ID NO: 10553, SEQ ID NO: 10555, SEQ ID NO: 10556, SEQ ID NO: 10557, SEQ ID NO: 10558, SEQ ID NO: 10560, SEQ ID NO: 10562, SEQ ID NO: 10564, SEQ ID NO: 10566, SEQ ID NO: 10567, SEQ ID NO: 10569, SEQ ID NO: 10571, SEQ ID NO: 10573, SEQ ID NO: 10574, SEQ ID NO: 10576, SEQ ID NO: 10578, SEQ ID NO: 10580, SEQ ID NO: 10582, SEQ ID NO: 10584, SEQ ID NO: 10586, SEQ ID NO: 10588, SEQ ID NO: 10590, SEQ ID NO: 10592, SEQ ID NO: 10594, SEQ ID NO: 10596, SEQ ID NO: 10597, SEQ ID NO: 10599, SEQ ID NO: 10601, SEQ ID NO: 10603, SEQ ID NO: 10604, SEQ ID NO: 10606, SEQ ID NO: 10607, SEQ ID NO: 10609, SEQ ID NO: 10611, SEQ ID NO: 10613, SEQ ID NO: 10614, SEQ ID NO: 10615, SEQ ID NO: 10616, SEQ ID NO: 10618, SEQ ID NO: 10619, SEQ ID NO: 10620, SEQ ID NO: 10622, SEQ ID NO: 10624, SEQ ID NO: 10625, SEQ ID NO: 10627, SEQ ID NO: 10629,
SEQ ID NO: 10630, SEQ ID NO: 10632, SEQ ID NO: 10633, SEQ ID NO: 10635, SEQ ID NO: 10637, SEQ ID NO: 10639, SEQ ID NO: 10641, SEQ ID NO: 10642, SEQ ID NO: 10644, SEQ ID NO: 10646, SEQ ID NO: 10647, SEQ ID NO: 10648, SEQ ID NO: 10649, SEQ ID NO: 10650, SEQ ID NO: 10652, SEQ ID NO: 10654, SEQ ID NO: 10655, SEQ ID NO: 10656, SEQ ID NO: 10658, SEQ ID NO: 10659, SEQ ID NO: 10661, SEQ ID NO: 10663, SEQ ID NO: 10665, SEQ ID NO: 10667, SEQ ID NO: 10669, SEQ ID NO: 10670, SEQ ID NO: 10671, SEQ ID NO: 10673, SEQ ID NO: 10674, SEQ ID NO: 10676, SEQ ID NO: 10678, SEQ ID NO: 10680, SEQ ID NO: 10682, SEQ ID NO: 10683, SEQ ID NO: 10685, SEQ ID NO: 10687, SEQ ID NO: 10689, SEQ ID NO: 10691, SEQ ID NO: 10693, SEQ ID NO: 10695, SEQ ID NO: 10696, SEQ ID NO: 10698, SEQ ID NO: 10700, SEQ ID NO: 10702, SEQ ID NO: 10704, SEQ ID NO: 10706, SEQ ID NO: 10708, SEQ ID NO: 10710, SEQ ID NO: 10711, SEQ ID NO: 10713, SEQ ID NO: 10715, SEQ ID NO: 10717, SEQ ID NO: 10718, SEQ ID NO: 10720, SEQ ID NO: 10722, SEQ ID NO: 10723, SEQ ID NO: 10725, SEQ ID NO: 10727, SEQ ID NO: 10728, SEQ ID NO: 10730, SEQ ID NO: 10732, SEQ ID NO: 10734, SEQ ID NO: 10736, SEQ ID NO: 10738, SEQ ID NO: 10740, SEQ ID NO: 10742, SEQ ID NO: 10744, SEQ ID NO: 10746, SEQ ID NO: 10748, SEQ ID NO: 10750, SEQ ID NO: 10752, SEQ ID NO: 10753, SEQ ID NO: 10754, SEQ ID NO: 10756, SEQ ID NO: 10757, SEQ ID NO: 10758, SEQ ID NO: 10760, SEQ ID NO: 10761, SEQ ID NO: 10763, SEQ ID NO: 10765, SEQ ID NO: 10767, SEQ ID NO: 10769, SEQ ID NO: 10771, SEQ ID NO: 10773, SEQ ID NO: 10774, SEQ ID NO: 10776, SEQ ID NO: 10778, SEQ ID NO: 10780, SEQ ID NO: 10781, SEQ ID NO: 10783,

NO: 12019, SEQ ID NO: 12020, SEQ ID NO: 12022, SEQ ID NO: 12024, SEQ ID NO: 12026, SEQ ID NO: 12028, SEQ ID NO: 12030, SEQ ID NO: 12032, SEQ ID NO: 12034, SEQ ID NO: 12036, SEQ ID NO: 12038, SEQ ID NO: 12040, SEQ ID NO: 12042, SEQ ID NO: 12044, SEQ ID NO: 12046, SEQ ID NO: 12048, SEQ ID NO: 12049, SEQ ID NO: 12051, SEQ ID NO: 12053, SEQ ID NO: 12055, SEQ ID NO: 12056, SEQ ID NO: 12058, SEQ ID NO: 12060, SEQ ID NO: 12062, SEQ ID NO: 12064, SEQ ID NO: 12066, SEQ ID NO: 12068, SEQ ID NO: 12070, SEQ ID NO: 12072, SEQ ID NO: 12074, SEQ ID NO: 12076, SEQ ID NO: 12078, SEQ ID NO: 12080,

SEQ ID NO: 12082, SEQ ID NO: 12084, SEQ ID NO: 12086, SEQ ID NO: 12088, SEQ ID NO: 12090, SEQ ID NO: 12092, SEQ ID NO: 12094, SEQ ID NO: 12096, SEQ ID NO: 12097, SEQ ID NO: 12099, SEQ ID NO: 12101, SEQ ID NO: 12103, SEQ ID NO: 12105, SEQ ID NO: 12107, SEQ ID NO: 12109, SEQ ID NO: 12111, SEQ ID NO: 12113, SEQ ID NO: 12115, SEQ ID NO: 12117, SEQ ID NO: 12119, SEQ ID NO: 12121, SEQ ID NO: 12123, SEQ ID NO: 12125, SEQ ID NO: 12127, SEQ ID NO: 12129, SEQ ID NO: 12131, SEQ ID NO: 12133, SEQ ID NO: 12135, SEQ ID NO: 12137, SEQ ID NO: 12139, SEQ ID NO: 12141, SEQ ID NO: 12143, SEQ ID NO: 12145, SEQ ID NO: 12147, SEQ ID NO: 12149, SEQ ID NO: 12151, SEQ ID NO: 12153, SEQ ID NO: 12155, SEQ ID NO: 12157, SEQ ID NO: 12159, SEQ ID NO: 12161, SEQ ID NO: 12163, SEQ ID NO: 12165, SEQ ID NO: 12167, SEQ ID NO: 12169, SEQ ID NO: 12171, SEQ ID NO: 12173, SEQ ID NO: 12175, SEQ ID NO: 12177, SEQ ID NO: 12179, SEQ ID NO: 12181, SEQ ID NO: 12183, SEQ ID NO: 12185, SEQ ID NO: 12187, SEQ ID NO: 12189, SEQ ID NO: 12191, SEQ ID NO: 12193, SEQ ID NO: 12195, SEQ ID NO: 12197, SEQ ID NO: 12199, SEQ ID NO: 12201, SEQ ID NO: 12203, SEQ ID NO: 12205, SEQ ID NO: 12207, SEQ ID NO: 12209, SEQ ID NO: 12211, SEQ ID NO: 12213, SEQ ID NO: 12215, SEQ ID NO: 12216, SEQ ID NO: 12218, SEQ ID NO: 12220, SEQ ID NO: 12222, SEQ ID NO: 12224, SEQ ID NO: 12226, SEQ ID NO: 12228, SEQ ID NO: 12230, SEQ ID NO: 12232, SEQ ID NO: 12234, SEQ ID NO: 12236, SEQ ID NO: 12238, SEQ ID NO: 12240, SEQ ID NO: 12242, SEQ ID NO: 12244, SEQ ID NO: 12246, SEQ ID NO: 12248, SEQ ID NO: 12250, SEQ ID NO: 12252, SEQ ID NO: 12254, SEQ ID NO: 12256, SEQ ID NO: 12258, SEQ ID NO: 12260, SEQ ID NO: 12262, SEQ ID NO: 12264, SEQ ID NO: 12266, SEQ ID NO: 12268, SEQ ID NO: 12270, SEQ ID NO: 12272, SEQ ID NO: 12274, SEQ ID NO: 12276, SEQ ID NO: 12278,

SEQ ID NO: 12279, SEQ ID NO: 12281, SEQ ID NO: 12283, SEQ ID NO: 12285, SEQ ID NO: 12287, SEQ ID NO: 12289, SEQ ID NO: 12291, SEQ ID NO: 12293, SEQ ID NO: 12295, SEQ ID NO: 12297, SEQ ID NO: 12299, SEQ ID NO: 12301, SEQ ID NO: 12303, SEQ ID NO: 12305, SEQ ID NO: 12307, SEQ ID NO: 12309, SEQ ID NO: 12311, SEQ ID NO: 12313, SEQ ID NO: 12315, SEQ ID NO: 12317, SEQ ID NO: 12319, SEQ ID NO: 12321, SEQ ID NO: 12323, SEQ ID NO: 12325, SEQ ID NO: 12327, SEQ ID NO: 12329, SEQ ID NO: 12331, SEQ ID NO: 12333, SEQ ID NO: 12335, SEQ ID NO: 12337, SEQ ID NO: 12339, SEQ ID NO: 12341, SEQ ID NO: 12343, SEQ ID NO: 12345, SEQ ID NO: 12347, SEQ ID NO: 12349, SEQ ID NO: 12351, SEQ ID NO: 12353, SEQ ID NO: 12354, SEQ ID NO: 12356, SEQ ID NO: 12358, SEQ ID NO: 12360, SEQ ID NO: 12362, SEQ ID NO: 12364, SEQ ID NO: 12366, SEQ ID NO: 12368, SEQ ID NO: 12370, SEQ ID NO: 12372, SEQ ID NO: 12374, SEQ ID NO: 12376, SEQ ID NO: 12378, SEQ ID NO: 12380, SEQ ID NO: 12382, SEQ ID NO: 12384, SEQ ID NO: 12386, SEQ ID NO: 12388, SEQ ID NO: 12390, SEQ ID NO: 12392, SEQ ID NO: 12394, SEQ ID NO: 12396, SEQ ID NO: 12398, SEQ ID NO: 12400, SEQ ID NO: 12402, SEQ ID NO: 12403, SEQ ID NO: 12405, SEQ ID NO: 12407, SEQ ID NO: 12409, SEQ ID NO: 12410, SEQ ID NO: 12412, SEQ ID NO: 12414, SEQ ID NO: 12416, SEQ ID NO: 12417, SEQ ID NO: 12419, SEQ ID NO: 12421, SEQ ID NO: 12423, SEQ ID NO: 12425, SEQ ID NO: 12427, SEQ ID NO: 12429, SEQ ID NO: 12431, SEQ ID NO: 12433, SEQ ID NO: 12435, SEQ ID NO: 12436, SEQ ID NO: 12438, SEQ ID NO: 12440, SEQ ID NO: 12442, SEQ ID NO: 12444, SEQ ID NO: 12446, SEQ ID NO: 12448, SEQ ID NO: 12450, SEQ ID NO: 12452, SEQ ID NO: 12454, SEQ ID NO: 12456, SEQ ID NO: 12458, SEQ ID NO: 12460, SEQ ID

SEQ ID NO: 18938, SEQ ID NO: 18940, SEQ ID NO: 18941, SEQ ID NO:-18943, SEQ ID NO: 18944, SEQ ID NO: 18946, SEQ ID NO: 18947, SEQ ID NO: 18949, SEQ ID NO: 18951, SEQ ID NO: 18953, SEQ ID NO: 18955, SEQ ID NO: 18956, SEQ ID NO: 18957, SEQ ID NO: 18958, SEQ ID NO: 18959, SEQ ID NO: 18960, SEQ ID NO: 18962, SEQ ID NO: 18964, SEQ ID NO: 18966, SEQ ID NO: 18968, SEQ ID NO: 18969, SEQ ID NO: 18970, SEQ ID NO: 18972, SEQ ID NO: 18973, SEQ ID NO: 18975, SEQ ID NO: 18976, SEQ ID NO: 18978, SEQ ID NO: 18980,
SEQ ID NO: 18981, SEQ ID NO: 18982, SEQ ID NO: 18983, SEQ ID NO: 18984, SEQ ID NO: 18985, SEQ ID NO: 18986, SEQ ID NO: 18987, SEQ ID NO: 18988, SEQ ID NO: 18989, SEQ ID NO: 18990, SEQ ID NO: 18992, SEQ ID NO: 18993, SEQ ID NO: 18995, SEQ ID NO: 18997, SEQ ID NO: 18998, SEQ ID NO: 18999, SEQ ID NO: 19000, SEQ ID NO: 19001, SEQ ID NO: 19002, SEQ ID NO: 19004, SEQ ID NO: 19006
SEQ ID NO: 19007, SEQ ID NO: 19009, SEQ ID NO: 19011, SEQ ID NO: 19012, SEQ ID NO: 19013, SEQ ID NO: 19014, SEQ ID NO: 19016, SEQ ID NO: 19018, SEQ ID NO: 19020, SEQ ID NO: 19022, SEQ ID NO: 19024, and SEQ ID NO: 19025

(b) a polynucleotide comprising a nucleotide sequence encoding a protein comprising the amino acid sequence set forth in any one of the following SEQ ID NOs:

SEQ ID NO:10469, SEQ ID NO:10474, SEQ ID NO:10476, SEQ ID NO:10478, SEQ ID NO:10480, SEQ ID NO:10482, SEQ ID NO:10484, SEQ ID NO:10486, SEQ ID NO:10490, SEQ ID NO:10492, SEQ ID NO:10494, SEQ ID NO:10499, SEQ ID NO:10501, SEQ ID NO:10506, SEQ ID NO:10509, SEQ ID NO:10513, SEQ ID NO:10515, SEQ ID NO:10518, SEQ ID NO:10520, SEQ ID NO:10522, SEQ ID NO:10525, SEQ ID NO:10527, SEQ ID NO:10531, SEQ ID NO:10533, SEQ ID NO:10536, SEQ ID NO:10538, SEQ ID NO:10541, SEQ ID NO:10544, SEQ ID NO:10547, SEQ ID NO:10549, SEQ ID NO:10552, SEQ ID NO:10554, SEQ ID NO:10559, SEQ ID NO:10561, SEQ ID NO:10563, SEQ ID NO:10565, SEQ ID NO:10568, SEQ ID NO:10570, SEQ ID NO:10572, SEQ ID NO:10575, SEQ ID NO:10577, SEQ ID NO:10579, SEQ ID NO:10581, SEQ ID NO:10583, SEQ ID NO:10585, SEQ ID NO:10587, SEQ ID NO:10589, SEQ ID NO:10591, SEQ ID NO:10593, SEQ ID NO:10595, SEQ ID NO:10598, SEQ ID NO:10600, SEQ ID NO:10602, SEQ ID NO:10605, SEQ ID NO:10608, SEQ ID NO:10610, SEQ ID NO:10612, SEQ ID NO:10617, SEQ ID NO:10621, SEQ ID NO:10623, SEQ ID NO:10626, SEQ ID NO:10628, SEQ ID NO:10631, SEQ ID NO:10634, SEQ ID NO:10636, SEQ ID NO:10638, SEQ ID NO:10640, SEQ ID NO:10643, SEQ ID NO:10645, SEQ ID NO:10651, SEQ ID NO:10653, SEQ ID NO:10657, SEQ ID NO:10660, SEQ ID NO:10662, SEQ ID NO:10664, SEQ ID NO:10666, SEQ ID NO:10668, SEQ ID NO:10672, SEQ ID NO:10675, SEQ ID NO:10677, SEQ ID NO:10679, SEQ ID NO:10681, SEQ ID NO:10684, SEQ ID NO:10686, SEQ ID NO:10688, SEQ ID NO:10690, SEQ ID NO:10692, SEQ ID NO:10694, SEQ ID NO:10697, SEQ ID NO:10699, SEQ ID NO:10701, SEQ ID NO:10703, SEQ ID NO:10705, SEQ ID NO:10707, SEQ ID NO:10709, SEQ ID NO:10712, SEQ ID NO:10714, SEQ ID NO:10716, SEQ ID NO:10719, SEQ ID NO:10721,
SEQ ID NO:10724, SEQ ID NO:10726, SEQ ID NO:10729, SEQ ID NO:10731, SEQ ID NO:10733, SEQ ID NO:10735, SEQ ID NO:10737, SEQ ID NO:10739, SEQ ID NO:10741, SEQ ID NO:10743, SEQ ID NO:10745, SEQ ID NO:10747, SEQ ID NO:10749, SEQ ID NO:10751, SEQ ID NO:10755, SEQ ID NO:10759, SEQ ID NO:10762, SEQ ID NO:10764, SEQ ID NO:10766, SEQ ID NO:10768, SEQ ID NO:10770, SEQ ID NO:10772, SEQ ID NO:10775, SEQ ID NO:10777, SEQ ID NO:10779, SEQ ID NO:10782, SEQ ID NO:10784, SEQ ID NO:10787, SEQ ID NO:10789, SEQ ID NO:10791, SEQ ID NO:10794, SEQ ID NO:10796, SEQ ID NO:10798, SEQ ID NO:10801, SEQ ID NO:10803, SEQ ID NO:10806, SEQ ID NO:10809, SEQ ID NO:10811, SEQ ID NO:10813, SEQ ID NO:10816, SEQ ID NO:10819, SEQ ID NO:10821, SEQ ID NO:10823, SEQ ID NO:10825, SEQ ID NO:10827, SEQ ID NO:10829, SEQ ID NO:10833, SEQ ID NO:10835, SEQ ID NO:10837, SEQ ID NO:10839, SEQ ID NO:10843, SEQ ID NO:10846, SEQ ID NO:10848, SEQ ID NO:10851,

SEQ ID NO:11836, SEQ ID NO:11838, SEQ ID NO:11840, SEQ ID NO:11842, SEQ ID NO:11844, SEQ ID NO:11846, SEQ ID NO:11848, SEQ ID NO:11852, SEQ ID NO:11854, SEQ ID NO:11857, SEQ ID NO:11860, SEQ ID NO:11862, SEQ ID NO:11864, SEQ ID NO:11866, SEQ ID NO:11868, SEQ ID NO:11872, SEQ ID NO:11874, SEQ ID NO:11876, SEQ ID NO:11878, SEQ ID NO:11881, SEQ ID NO:11883, SEQ ID NO:11885, SEQ ID NO:11887, SEQ ID NO:11889, SEQ ID NO:11891, SEQ ID NO:11894, SEQ ID NO:11896, SEQ ID NO:11898, SEQ ID NO:11900, SEQ ID NO:11902, SEQ ID NO:11904, SEQ ID NO:11906, SEQ ID NO:11908, SEQ ID NO:11910, SEQ ID NO:11912, SEQ ID NO:11914, SEQ ID NO:11917, SEQ ID NO:11920, SEQ ID NO:11922, SEQ ID NO:11924, SEQ ID NO:11926, SEQ ID NO:11929, SEQ ID NO:11931, SEQ ID NO:11933, SEQ ID NO:11935, SEQ ID NO:11937, SEQ ID NO:11939, SEQ ID NO:11941, SEQ ID NO:11943, SEQ ID NO:11945, SEQ ID NO:11947, SEQ ID NO:11950, SEQ ID NO:11952, SEQ ID NO:11954, SEQ ID NO:11956, SEQ ID NO:11958, SEQ ID NO:11960, SEQ ID NO:11962, SEQ ID NO:11964, SEQ ID NO:11966, SEQ ID NO:11968, SEQ ID NO:11970, SEQ ID NO:11972, SEQ ID NO:11974, SEQ ID NO:11976, SEQ ID NO:11979, SEQ ID NO:11981, SEQ ID NO:11983, SEQ ID NO:11985, SEQ ID NO:11987, SEQ ID NO:11989, SEQ ID NO:11991, SEQ ID NO:11993, SEQ ID NO:11995, SEQ ID NO:11997, SEQ ID NO:11999, SEQ ID NO:12001, SEQ ID NO:12004, SEQ ID NO:12006, SEQ ID NO:12009, SEQ ID NO:12011, SEQ ID NO:12013, SEQ ID NO:12015, SEQ ID NO:12017, SEQ ID NO:12021, SEQ ID NO:12023, SEQ ID NO:12025, SEQ ID NO:12027, SEQ ID NO:12029, SEQ ID NO:12031, SEQ ID NO:12033, SEQ ID NO:12035, SEQ ID NO:12037, SEQ ID NO:12039, SEQ ID NO:12041, SEQ ID NO:12043, SEQ ID NO:12045, SEQ ID NO:12047, SEQ ID NO:12050, SEQ ID NO:12052,
SEQ ID NO:12054, SEQ ID NO:12057, SEQ ID NO:12059, SEQ ID NO:12061, SEQ ID NO:12063, SEQ ID NO:12065, SEQ ID NO:12067, SEQ ID NO:12069, SEQ ID NO:12071, SEQ ID NO:12073, SEQ ID NO:12075, SEQ ID NO:12077, SEQ ID NO:12079, SEQ ID NO:12081, SEQ ID NO:12083, SEQ ID NO:12085, SEQ ID NO:12087, SEQ ID NO:12089, SEQ ID NO:12091, SEQ ID NO:12093, SEQ ID NO:12095, SEQ ID NO:12098, SEQ ID NO:12100, SEQ ID NO:12102, SEQ ID NO:12104, SEQ ID NO:12106, SEQ ID NO:12108, SEQ ID NO:12110, SEQ ID NO:12112, SEQ ID NO:12114, SEQ ID NO:12116, SEQ ID NO:12118, SEQ ID NO:12120, SEQ ID NO:12122, SEQ ID NO:12124, SEQ ID NO:12126, SEQ ID NO:12128, SEQ ID NO:12130, SEQ ID NO:12132, SEQ ID NO:12134, SEQ ID NO:12136, SEQ ID NO:12138, SEQ ID NO:12140, SEQ ID NO:12142, SEQ ID NO:12144, SEQ ID NO:12146, SEQ ID NO:12148, SEQ ID NO:12150, SEQ ID NO:12152, SEQ ID NO:12154, SEQ ID NO:12156, SEQ ID NO:12158, SEQ ID NO:12160, SEQ ID NO:12162, SEQ ID NO:12164, SEQ ID NO:12166, SEQ ID NO:12168, SEQ ID NO:12170, SEQ ID NO:12172, SEQ ID NO:12174, SEQ ID NO:12176, SEQ ID NO:12178, SEQ ID NO:12180, SEQ ID NO:12182, SEQ ID NO:12184, SEQ ID NO:12186, SEQ ID NO:12188, SEQ ID NO:12190, SEQ ID NO:12192, SEQ ID NO:12194, SEQ ID NO:12196, SEQ ID NO:12198, SEQ ID NO:12200, SEQ ID NO:12202, SEQ ID NO:12204, SEQ ID NO:12206, SEQ ID NO:12208, SEQ ID NO:12210, SEQ ID NO:12212, SEQ ID NO:12214, SEQ ID NO:12217, SEQ ID NO:12219, SEQ ID NO:12221, SEQ ID NO:12223, SEQ ID NO:12225, SEQ ID NO:12227, SEQ ID NO:12229, SEQ ID NO:12231, SEQ ID NO:12233, SEQ ID NO:12235, SEQ ID NO:12237, SEQ ID NO:12239, SEQ ID NO:12241, SEQ ID NO:12243, SEQ ID NO:12245, SEQ ID NO:12247, SEQ ID NO:12249, SEQ ID NO:12251, SEQ ID NO:12253, SEQ ID NO:12255,
SEQ ID NO:12257, SEQ ID NO:12259, SEQ ID NO:12261, SEQ ID NO:12263, SEQ ID NO:12265, SEQ ID NO:12267, SEQ ID NO:12269, SEQ ID NO:12271, SEQ ID NO:12273, SEQ ID NO:12275, SEQ ID NO:12277, SEQ ID NO:12280, SEQ ID NO:12282, SEQ ID NO:12284, SEQ ID NO:12286, SEQ ID NO:12288, SEQ ID NO:12290, SEQ ID NO:12292, SEQ ID NO:12294, SEQ ID NO:12296, SEQ ID NO:12298, SEQ ID NO:12300, SEQ ID NO:12302, SEQ ID NO:12304, SEQ ID NO:12306, SEQ ID NO:12308, SEQ ID NO:12310,

NO:18662, SEQ ID NO:18665, SEQ ID NO:18668, SEQ ID NO:18672, SEQ ID NO:18674, SEQ ID NO:18676, SEQ ID NO:18678, SEQ ID NO:18680, SEQ ID NO:18682, SEQ ID NO:18686, SEQ ID NO:18688, SEQ ID NO:18691, SEQ ID NO:18693, SEQ ID NO:18695, SEQ ID NO:18697, SEQ ID NO:18699, SEQ ID NO:18701, SEQ ID NO:18703, SEQ ID NO:18705, SEQ ID NO:18709, SEQ ID NO:18711, SEQ ID NO:18713, SEQ ID NO:18715, SEQ ID NO:18717, SEQ ID NO:18720, SEQ ID NO:18722, SEQ ID NO:18724, SEQ ID NO:18726, SEQ ID NO:18729, SEQ ID NO:18731, SEQ ID NO:18733, SEQ ID NO:18735, SEQ ID NO:18737, SEQ ID NO:18739, SEQ ID NO:18741, SEQ ID NO:18743, SEQ ID NO:18745, SEQ ID NO:18747, SEQ ID NO:18749, SEQ ID NO:18751, SEQ ID NO:18753, SEQ ID NO:18759, SEQ ID NO:18763, SEQ ID NO:18765, SEQ ID NO:18770, SEQ ID NO:18773, SEQ ID NO:18775, SEQ ID NO:18777, SEQ ID NO:18779, SEQ ID NO:18781, SEQ ID NO:18783, SEQ ID NO:18785, SEQ ID NO:18787, SEQ ID NO:18790, SEQ ID NO:18793, SEQ ID NO:18795, SEQ ID NO:18797, SEQ ID NO:18800, SEQ ID NO:18802, SEQ ID NO:18804, SEQ ID NO:18806, SEQ ID NO:18809, SEQ ID NO:18811, SEQ ID NO:18813, SEQ ID NO:18815, SEQ ID NO:18817, SEQ ID NO:18819, SEQ ID NO:18822, SEQ ID NO:18824, SEQ ID NO:18826, SEQ ID NO:18828, SEQ ID NO:18830, SEQ ID NO:18832, SEQ ID NO:18834, SEQ ID NO:18836, SEQ ID NO:18840, SEQ ID NO:18843, SEQ ID NO:18847, SEQ ID NO:18850, SEQ ID NO:18853, SEQ ID NO:18855, SEQ ID NO:18857, SEQ ID NO:18859, SEQ ID NO:18862, SEQ ID NO:18865, SEQ ID NO:18868, SEQ ID NO:18870,
SEQ ID NO:18874, SEQ ID NO:18876, SEQ ID NO:18879, SEQ ID NO:18882, SEQ ID NO:18884, SEQ ID NO:18888, SEQ ID NO:18891, SEQ ID NO:18894, SEQ ID NO:18896, SEQ ID NO:18898, SEQ ID NO:18900, SEQ ID NO:18902, SEQ ID NO:18906, SEQ ID NO:18908, SEQ ID NO:18910, SEQ ID NO:18912, SEQ ID NO:18914, SEQ ID NO:18916, SEQ ID NO:18918, SEQ ID NO:18920, SEQ ID NO:18922, SEQ ID NO:18924, SEQ ID NO:18926, SEQ ID NO:18929, SEQ ID NO:18931, SEQ ID NO:18933, SEQ ID NO:18935, SEQ ID NO:18937, SEQ ID NO:18939, SEQ ID NO:18942, SEQ ID NO:18945, SEQ ID NO:18948, SEQ ID NO:18950, SEQ ID NO:18952, SEQ ID NO:18954, SEQ ID NO:18961, SEQ ID NO:18963, SEQ ID NO:18965, SEQ ID NO:18967, SEQ ID NO:18971, SEQ ID NO:18974, SEQ ID NO:18977, SEQ ID NO:18979,
SEQ ID NO:18991, SEQ ID NO:18994, SEQ ID NO:18996, SEQ ID NO:19003, SEQ ID NO:19005
SEQ ID NO:19008, SEQ ID NO:19010, SEQ ID NO:19015, SEQ ID NO:19017, SEQ ID NO:19019, SEQ ID NO:19021, and SEQ ID NO:19023

(c) a polynucleotide comprising a nucleotide sequence encoding a protein comprising an amino acid sequence selected from the amino acid sequences of (b), in which one or more amino acids are substituted, deleted, inserted, and/or added, wherein said protein is functionally equivalent to the protein comprising said amino acid sequence selected from the amino acid sequences of (b);
(d) a polynucleotide that hybridizes with a polynucleotide comprising a nucleotide sequence selected from the nucleotide sequences of (a), and that comprises a nucleotide sequence encoding a protein functionally equivalent to the protein encoded by the nucleotide sequence selected from the nucleotide sequences of (a);
(e) a polynucleotide comprising a nucleotide sequence encoding a partial amino acid sequence of a protein encoded by the polynucleotide of (a) to (d);
(f) a polynucleotide comprising a nucleotide sequence with at least 70% identity to the nucleotide sequence of (a).

9. A substantially pure protein encoded by the polynucleotide of claim 8.

10. An antibody against the protein or peptide of any one of claims 6, 7, and 9.